(Founded in 1935 by Carl Murchison)

The Journal of PSYCHOLOGY

The General Field of Psychology

POWELL MURCHISON
Managing Editor

HELEN M. DAVIS
Associate Editor

Editorial Board

JOHN A. ANTOINETTI
Life Insurance Agency
Management Association

ARTHUR G. BILLS
University of Cincinnati

E. A. BOTT
University of Toronto

LEONARD CARMICHAEL
National Geographic Society

HERBERT S. CONRAD
U. S. Office of Education

JOHN F. DASHIELL
Florida State University

LEONARD W. DOOB
Yale University

ROY M. DORCUS
University of California
at Los Angeles

FRANK A. GELDARD
Princeton University

CLARENCE H. GRAHAM
Columbia University

ERNEST R. HILGARD
Stanford University

JOHN E. HORROCKS
Ohio State University

HEINRICH KLÜVER
University of Chicago

NORMAN L. MUNN
University of Adelaide,
Australia

WARREN H. TEICHNER
Northeastern University

LEE EDWARD TRAVIS
University of Southern
California

ROBERT C. TRYON
University of California

JOSEPH ZUBIN
Psychiatric Institute,
Columbia University

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from date of publication.

VOLUME 72

1969

Copyright © 1969 by The Journal Press
Provincetown, Massachusetts, U. S. A.
All rights reserved

Published at 35 New St., Worcester, Mass.
Send all mail to Managing Editor, 2 Commercial Street, Provincetown, Massachusetts 02657

AUTHOR INDEX

Rosalie K. Ambler 263
Herbert C. Archibald 151
Patricia Asta 159
William C. Bingham 247
Robert T. Bodensteiner 165
Susan H. Borcyck 211
John R. Braun 159
Henry R. Burke 247
Kenneth A. Chandler 219
Virginia C. Clark 21
Lora-Jean Collett 179
Angelo Danesino 71
Carolyn Draeger 11
Francis M. Dwyer, Jr. 61
Eugene S. Edgington 143
Gerald Engel 135
Donald W. Felker 131
Diane Fiorentino 93
John F. Fracchia 93
Kentner V. Fritz 115
Martin F. Fritz 115
Malcolm Gynther 165
C. Wade Harrison 125
Brenton Hazlewood 195
B. Kent Houston 65
Jean S. Kerrick 21
Vernon L. Kiker, Jr. 41
Dennis Kneff 165
Earle A. Koile 11
Alan Krichev 195
John J. LaGaipa 109
Gerald Larson 131
William A. Layman 71
David Lester 179
M. S. Lindauer 3
Sven Lundstedt 85
Sydney Merlis 93
Frank Milhollan 131
Bishwa Nath Mukherjee 119
Judith W. Onley 211
David E. Parker 151
Gary W. R. Patton 199
Darhl M. Pedersen 101
Joseph G. Phelan 41
John R. Pyper 131
Donna J. Rawls 125
James R. Rawls 125
Robert H. Reiher 41
Charles E. Schaefer 233
Eduard Schludermann 33, 183, 243, 253
Shirin Schludermann 33, 183, 243, 253
Rio Sciortino 169
Charles Sheppard 93
Joyce L. Sichel 219
Kay Smith 101
M. Kent Smith 151
Gordon Stanley 99
Nina Sullivan 211
H. Edwin Titus 35
E. Paul Torrance 189
Read D. Tuddenham 151
Thomas W. Vodde 165
John Voevodsky 269
Fred W. Vondracek 55
L. K. Waters 263
Saundra F. Zimmerman 101

TABLE OF CONTENTS

Quantitative analysis of psychoanalytic studies of Shakespeare . . . 3
By M. S. Lindauer

T-group member ratings of leader and self in a human relations laboratory . 11
By Earle A. Koile and Carolyn Draeger

The relationship of illness and health behavior to paper-and-pencil tests . . 21
By Jean S. Kerrick and Virginia A. Clark

A note on the use of discrepancy scores in a self-concept inventory . . 33
By Shirin Schludermann and Eduard Schludermann

Prediction of supervisory success by use of standard psychological tests . . . 35
By H. Edwin Titus

Reversal and nonreversal shifts in neurologically handicapped children . . 41
By Robert H. Reiher, Joseph G. Phelan, and Vernon L. Kiker, Jr.

The study of self-disclosure in experimental interviews 55
By Fred W. Vondracek

An analysis of the instructional effectiveness of visual illustrations presented via television 61
By Francis M. Dwyer, Jr.

Field independence and performance in distraction 65
By B. Kent Houston

Contrasting personality patterns of high and low achievers among college students of Italian and Irish descent 71
By Angelo Danesino and William A. Layman

Explorations in methodology: The scientific question as a schemapiric dimension 85
By Sven Lundstedt

A comparison of techniques for the scoring of avoidable errors on the Raven Progressive Matrices 93
By John F. Fracchia, Diane Fiorentino, Charles Sheppard, and Sydney Merlis

Laymen's identification of item content on the Eysenck Personality Inventory . 99
By Gordon Stanley

The factorial determination of types of conforming individuals . . . 101
By Saundra F. Zimmerman, Darhl M. Pedersen, and Kay Smith

Biographical inventories and style of leadership 109
By John J. La Gaipa

Untimed norms for the SRA Mechanical Knowledge and Shop Arithmetic Tests 115
By Kentner V. Fritz and Martin F. Fritz

The role of verbal factor in a forced-choice test of achievement value . . 119
By Bishwa Nath Mukherjee

An investigation of success predictors in graduate school in psychology . . 125
By James R. Rawls, Donna J. Rawls, and C. Wade Harrison

The effect of sample items on test performance 131
By Donald W. Felker, Gerald Larson, Frank Milhollan, and John R. Pyper

Comparison between American permanent residents of Israel: Part II, Israeli background 135
By Gerald Engel

Approximate randomization tests 143
By Eugene S. Edgington

Who seeks treatment? A study of psychiatrically disabled veterans . . . 151
By Herbert C. Archibald, David E. Parker, M. Kent Smith, and Read D. Tuddenham

A comparison of "real" vs. "ideal" self with a self-actualization inventory . 159
By John R. Braun and Patricia Asta

The clinical utility of Block's ego-resiliency and ego-control scales . . . 165
By Dennis Kneff, Robert T. Bodensteiner, Thomas W. Vodde, and Malcolm Gynther

Factorial study and analysis of factor variance of general adaptability self-ratings from a combined sample of male and female subjects . . . 169
By Rio Sciortino

The fear of death and the fear of dying 179
By Lora-Jean Collett and David Lester

Social role perceptions of children in Hutterite communal society . . . 183
By Shirin Schludermann and Eduard Schludermann

Peer influences on preschool children's willingness to try difficult tasks . . 189
By E. Paul Torrance

The use of matrix rather than serial presentations of nonsense syllables . . 195
By Alan Krichev and Brenton Hazlewood

Interstressor consistency of subject differences in autonomic activity . . 199
By Gary W. R. Patton

Magnitude scaling of foods 211
By Judith W. Onley, Nina Sullivan, and Susan H. Borcyck

The color-word interference test: The effects of varied color-word combinations upon verbal response latency 219
By Joyce L. Sichel and Kenneth A. Chandler

The self-concept of creative adolescents 233
By Charles E. Schaefer

Developmental study of social role perception among Hutterite adolescents . 243
By Shirin Schludermann and Eduard Schludermann

Raven's Progressive Matrices: More on construct validity 247
By Henry R. Burke and William C. Bingham

Scale checking style as a function of age and sex in Indian and Hutterite children 253
By Shirin Schludermann and Eduard Schludermann

A comparison of volunteers and nonvolunteers for astronaut training . . 263
By L. K. Waters and Rosalie K. Ambler

Quantitative behavior of warring nations 269
By John Voevodsky

Journal of Genetic Psychology

Founded by G. Stanley Hall in 1891; reorganized by Carl Murchison in 1925

Quarterly, beginning with March. Devoted to child behavior, animal behavior, and the problems of aging. Manuscripts may be sent to any member of the Editorial Board or directly to The Journal Press. Subscriptions should be sent directly to The Journal Press, and business communications addressed to the Managing Editor. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1891 at $15.00 per volume, plus transportation.

Journal of General Psychology

Founded in 1927 by E. B. Titchener and Carl Murchison

Quarterly, beginning with January. Devoted primarily to experimental, theoretical, clinical, and historical psychology. Manuscripts may be sent to any member of the Editorial Board or directly to The Journal Press. Subscriptions should be sent directly to The Journal Press, and business communications addressed to the Managing Editor. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1928 at $15.00 per volume, plus transportation.

Journal of Social Psychology

Founded in 1929 by John Dewey and Carl Murchison

Bimonthly, beginning with February. Devoted to studies of persons in group settings and of culture and personality; special attention to cross-cultural articles and notes, and to briefly reported replications and refinements. Manuscripts may be sent to any member of the Editorial Board or directly to The Journal Press. Subscriptions should be sent directly to The Journal Press, and business communications addressed to the Managing Editor. Three volumes annually. Per annum, $30.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1930 at $15.00 per volume, plus transportation.

Each author will receive 100 unbound reprints gratis. Additional reprints may be purchased at a nominal cost when the author returns page-proof.

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts 02657
U. S. A.

Club Rate for Five International Psychological Journals

	PER ANNUM
THE JOURNAL OF GENETIC PSYCHOLOGY (Quarterly, two volumes per year)	$20.00
GENETIC PSYCHOLOGY MONOGRAPHS (Quarterly, two volumes per year)	$20.00
THE JOURNAL OF GENERAL PSYCHOLOGY (Quarterly, two volumes per year)	$20.00
THE JOURNAL OF SOCIAL PSYCHOLOGY (Bimonthly, three volumes per year)	$30.00
THE JOURNAL OF PSYCHOLOGY (Bimonthly, three volumes per year)	$30.00
TOTAL	$120.00
TEACHERS AND STUDENTS EDITION, *Restricted Sale*	$50.00

The *Teachers and Students Edition* is open to any individual psychologist, teacher, student, or physician but requires an agreement that the edition will not be offered for sale till after five years from publication and an agreement that the edition will not be substituted for an institutional subscription.

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts 02657
U. S. A.

Genetic Psychology Monographs

Founded in 1925 by Carl Murchison

Quarterly, beginning with February. Devoted to child behavior, animal behavior, and the problems of aging. Each number one or more complete researches. Manuscripts may be sent to any member of the Editorial Board or directly to The Journal Press. Subscriptions should be sent directly to The Journal Press, and business communications addressed to the Managing Editor. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1926 at $15.00 per volume, plus transportation.

NOTICE. Effective immediately we offer prior publication for certain selected monographs. Charges will be based on the announced rates for *The Journal of Psychology. For a limited period of time our policy is changed for regular monographs: Authors will be charged for cuts and expensive tables, but no page charge will be imposed.* Two hundred reprints of each monograph will be supplied gratis.

The listing of monographs for 1926 through 1963 will be printed in the first half of each volume of all five journals. Monographs published after 1963 will be listed in the second half of each volume.

Genetic Psychology Monographs *(continued)*

VOLUME 1—1926

1. Performance tests for children of pre-school age—R. Stutsman
2. An experimental study of the eidetic type—H. Klüver

3 & 4. A study of natio-racial mental differences—N. D. M. Hirsch

5. A psychological study of juvenile delinquency by group methods—J. W. Bridges and K. M. B. Bridges
6. The influence of puberty praecox upon mental growth—A. Gesell

VOLUME 2—1927

1 & 2. The mind of a gorilla—R. M. Yerkes

3. The role of eye-muscles and mouth-muscles in the expression of the emotions—K. Dunlap
4. Family similarities in mental-test abilities—R. R. Willoughby
5. Coordination in the locomotion of infants—L. H. Burnside
6. The mind of a gorilla: Part II. Mental development—R. M. Yerkes

VOLUME 3—January-June, 1928

1. An experimental study of the olfactory sensitivity of the white rat—J. R. Liggett
2. A photographic study of eye movements in reading formulae—M. A. Tinker
3. An experimental study of the East Kentucky mountaineers—N. D. M. Hirsch
4. Responses of foetal guinea pigs prematurely delivered—G. T. Avery
5. Objective differentiation between three groups in education (teachers, research workers, and administrators)—M. B. Jensen
6. The effect of segregation on the sex behavior of the white rat as measured by the obstruction method—M. Jenkins

VOLUME 4—July-December, 1928

1. Observation and training of fundamental habits in young children—E. A. Bott, W. E. Blatz, N. Chant, and H. Bott

2 & 3. Determination of a content of the course in literature of a suitable difficulty for junior and senior high school students—M. C. Burch

4 & 5. Methods for diagnosis and treatment of cases of reading disability—M. Monroe

6. The relative effectiveness of lecture and individual reading as methods of college teaching—E. B. Greene

VOLUME 5—January-June, 1929

1. The age factor in animal learning: I. Rats in the problem box and the maze—C. P. Stone
2. The effect of delayed incentive on the hunger drive in the white rat—E. L. Hamilton
3. Which hand is the eye of the blind?—J. M. Smith
4. The effect of attitude on free word association-time—A. G. Ekdahl
5. The localization of tactual space: A study of average and constant errors under different types of localization—L. E. Cole
6. The effects of gonadectomy, vasotomy, and injections of placental and orchic extracts on the sex behavior of the white rat—H. W. Nissen

VOLUME 6—July-December, 1929

1. Learning and growth in identical infant twins: An experimental study by the method of co-twin control—A. Gesell and H. Thompson
2. The age factor in animal learning: II. Rats on a multiple light discrimination box and a difficult maze—C. P. Stone
3. The acquisition and interference of motor habits in young children—E. McGinnis
4. A vocational and socio-educational survey of graduates and non-graduates of small high schools of New England—A. D. Mueller

5 & 6. A study of the smiling and laughing of infants in the first year of life—R. W. Washburn

VOLUME 7—January-June, 1930

1. Tensions and emotional factors in reaction—E. Duffy
2. Teacher influence on class achievement: A study of the relationship of estimated teaching ability to pupil achievement in reading and arithmetic—H. R. Taylor

3 & 4. A study of the effect of inverted retinal stimulation upon spatially coordinated behavior—P. H. Ewert

5. A study of the mental development of children with lesion in the central nervous system—E. E. Lord
6. An experimental study upon three hundred school children over a six-year period—N. D. M. Hirsch

VOLUME 8—July-December, 1930

1. The amount and nature of activities of newborn infants under constant external stimulating conditions during the first ten days of life—O. C. Irwin
2. Race and social differences in performance tests—S. D. Porteus, *et al.*
3. Language and growth: The relative efficacy of early and deferred vocabulary training, studied by the method of co-twin control—L. C. Strayer
4. Eye-movements and optic nystagmus in early infancy—J. M. McGinnis

5 & 6. Reactions of kindergarten, first-, and second-grade children to constructive play materials—L. Farwell

VOLUME 9—January-June, 1931

1 & 2. The status of the first-born with special reference to intelligence—H. H. Hsiao

3 & 4. An experimental study of bright, average, and dull children at the four-year mental level—H. P. Davidson

5. An historical, critical, and experimental study of the Seashore-Kwalwasser test battery—P. R. Farnsworth
6. A comparison of difficulty and improvement in the learning of bright and dull children in reproducing a descriptive selection—F. T. Wilson

VOLUME 10—July-December, 1931

1. A comparative study of a group of southern white and negro infants—M. B. McGraw

2 & 3. An experimental study of prehension in infants by means of systematic cinema records—H. M. Halverson

4. The limits of learning ability in kittens—A. M. Shuey

5 & 6. The effect of habit interference upon performance in maze learning—O. W. Alm

VOLUME 11—January-June, 1932

1. General factors in transfer of training in the white rat—T. A. Jackson
2. The effect of color on visual apprehension and perception—M. A. Tinker
3. The reliability and validity of maze experiments with white rats—R. Leeper
4. A critical study of two lists of best books for children—F. K. Shuttleworth

5 & 6. Measuring human energy cost in industry: A general guide to the literature—R. M. Page

VOLUME 12—July-December, 1932

1. Family resemblances in verbal and numerical abilities—H. D. Carter
2. The development of fine prehension in infancy—B. M. Castner

3 & 4. The growth of adaptive behavior in infants: An experimental study at seven age levels—H. M. Richardson

5 & 6. Differential reactions to taste and temperature stimuli in newborn infants—K. Jensen

VOLUME 13—January-June, 1933

1. A critique of sublimation in males: A study of forty superior single men—W. S. Taylor
2. A study of the nature, measurement, and determination of hand preference—H. L. Koch, *et al.*
3. The growth and decline of intelligence: A study of a homogeneous group between the ages of ten and sixty—H. E. Jones and H. S. Conrad
4. The relation between the complexity of the habit to be acquired and the form of the learning curve in young children—M. L. Mattson
5. Eating habits in relation to personality development of two- and three-year-old children: A study of sixty-nine children in two nursery schools—A. A. Eliot
6. Coordinating mechanisms of the spinal cord—O. C. Ingebritsen

Genetic Psychology Monographs *(continued)*

VOLUME 14—July-December, 1933

Mental growth during the first three years: A developmental study of sixty-one children by repeated tests—N. Bayley
A study of triplets: including theories of their possible genetic relationships—F. N. Anderson and N. V. Scheidemann
The objective measurement of emotional reactions—H. V. Gaskill
Development of behavior in the fetal cat—J. D. Coronios
5. A study of certain language developments of children in grades four to twelve, inclusive—L. L. LaBrant
6. The effect of early and delayed practice on memory and motor performances studied by the method of co-twin control—J. R. Hilgard

VOLUME 15—January-June, 1934

1. Studies in the psychology of tone and music—P. R. Farnsworth
2. Motor learning of children in equilibrium in relation to nutrition—E. L. Beebe
3. Discrimination limens of pattern and size in the goldfish *Carassius auratus*—J. B. Rowley
4. Limits of learning ability in the white rat and the guinea pig—B. F. Riess

5 & 6. The limits of learning ability in rhesus monkeys—H. A. Fjeld

VOLUME 16—July-December, 1934

1. A statistical study of ratings on the California Behavior Inventory for Nursery-School Children—H. S. Conrad
2. An eye-movement study of objective examination questions—A. Frandsen
3. An experimental study of constitutional types—O. Klineberg, S. E. Asch, and H. Block
4. The development of a battery of objective group tests of manual laterality, with the results of their application to 1300 children—W. N. Durost

5 & 6. An experimental study in the prenatal guinea-pig of the origin and development of reflexes and patterns of behavior in relation to the stimulation of specific receptor areas during the period of active fetal life—L. Carmichael

VOLUME 17—January-December, 1935

1. Organization of behavior in the albino rat—R. L. Thorndike
2. Brightness discrimination in the rhesus monkey—M. P. Crawford
3. The limits of learning ability in cebus monkeys—A. M. Koch
4. Nature-nurture and intelligence—A. M. Leahy
5. On intelligence of epileptic children—E. B. Sullivan and L. Gahagan
6. A study of the play of children of preschool age by an unobserved observer—D. L. Cockrell

VOLUME 18—January-December, 1936

1. Sex differences in variational tendency—Q. McNemar and L. M. Terman
2. The process of learning to dress among nursery-school children—C. B. Key, M. R. White, M. P. Honzik, A. B. Heiney, and D. Erwin
3. A study of the present social status of a group of adults, who, when they were in elementary schools, were classified as mentally deficient—W. R. Baller
4. The influence of specific experience upon mental organizations—A. Anastasi

5 & 6. Studies in aggressiveness—L. Bender, S. Keiser, and P. Schilder

VOLUME 19—January-December, 1937

1. Psychological bases of self-mutilation—C. Dabrowski
Masculine temperament and secondary sex characteristics: A study of the relationship between psychological and physical measures of masculinity—H. Gilkinson
2. A psychological study of forty unmarried mothers—R. D. Nottingham
Behavior problems in the children of psychotic and criminal parents—L. Bender
3. Domination and integration in the social behavior of young children in an experimental play situation—H. H. Anderson
4. The sequential patterning of prone progression in the human infant—L. B. Ames

VOLUME 20—January-December, 1938

1. The relationship between characteristics of personality and physique in adolescents—P. S. de Q. Cabot
2. Behavior problems of elementary school children: A descriptive and comparative study—I. Y. Masten
Graphic representation of a man by four-year-old children in nine prescribed drawing situations—P. F. Gridley
3. Differences between two groups of adult criminals—R. S. Tolman
4. A comparative study by means of the Rorschach method of personality development in twenty pairs of identical twins—E. Troup
Individual differences in the facial expressive behavior of preschool children: A study by the time-sampling method—C. Swan

VOLUME 21—January-December, 1939

1. An experimental analysis of "level of aspiration"—R. Gould
2. Some light on the problem of bilingualism as found from a study of the progress in mastery of English among preschool children of non-American ancestry in Hawaii—M. E. Smith
3. Domination and social integration in the behavior of kindergarten children and teachers—H. H. Anderson
The capacity of the rhesus and cebus monkey and the gibbon to acquire differential response to complex visual stimuli—W. E. Galt
4. The social-sex development of children—E. H. Campbell

VOLUME 22—January-December, 1940

1. Measuring human relations: An introduction to the study of the interaction of individuals—E. D. Chapple
2. Aggressive behavior in young children and children's attitudes toward aggression—M. D. Fite
3. Student attitudes toward religion—E. Nelson
The prediction of the outcome-on-furlough of dementia praecox patients—J. S. Jacob
Significant characteristics of preschool children as located in the Conrad inventory—K. H. Read
4. Learning by children at noon-meal in a nursery school: Ten "good" eaters and ten "poor" eaters—J. B. McCay, E. B. Waring, and P. J. Kruse
Studies in the interpretation of play: I. Clinical observation of play disruption in young children—E. H. Erikson

VOLUME 23—January-June, 1941

1. An analysis of certain variables in a developmental study of language—F. M. Young
Infant development under conditions of restricted practice and of minimum social stimulation—W. Dennis
An analysis of the mental factors of various age groups from nine to sixty—B. Balinsky
2. Factors influencing performance on group and individual tests of intelligence: I. Rate of work—M. W. Bennett
Individual differences in apperceptive reaction: A study of the response of preschool children to pictures—E. W. Amen

VOLUME 24—July-December, 1941

1. Twins *T* and *C* from infancy to adolescence: A biogenetic study of individual differences by the method of co-twin control—A. Gesell and H. Thompson
Finger nail-biting: Its incipiency, incidence, and amelioration—A. L. Billig
An experimental study of the factors of maturation and practice in the behavioral development of the embryo of the frog, *Rana pipiens*—A. Fromme
2. The Fels child behavior scales—T. W. Richards and M. P. Simons
Measurement of the size of general English vocabulary through the elementary grades and high school—M. K. Smith
Stereotypes in the field of musical eminence—P. R. Farnsworth

VOLUME 25—January-June, 1942

1. A study of factors determining family size in a selected professional group—J. C. Flanagan
A genetic study of geometrical-optical illusions—A. Walters
Interpretation of behavior-ratings in terms of favorable and unfavorable deviations: A study of scores from the Read-Conrad Behavior Inventory—K. H. Read and H. S. Conrad
2. Are there any innate behavior tendencies?—J. B. Schoolland
An investigation of the intelligibility of the speech of the deaf—C. V. Hudgins and F. C. Numbers

Genetic Psychology Monographs *(continued)*

VOLUME 26—July-December, 1942

1. The critical frequency limen for visual flicker in children between the ages of 6 and 18—V. L. Miller
Some factors determining handedness in the white rat—K. L. Wentworth
2. Motivation and behavior—E. Frenkel-Brunswik

VOLUME 27—January-June, 1943

1. Comparison of children's personality traits, attitudes, and intelligence with parental occupation—N. R. Maddy
2. A comparative study of mental functioning patterns of problem and non-problem children seven, eight, and nine years of age—M. L. Pignatelli

VOLUME 28—July-December, 1943

1. Separation anxiety in young children: A study of hospital cases—H. Edelston
2. Correlates of vocational preferences—W. A. Bradley, Jr.

VOLUME 29—January-June, 1944

1. Mental changes after bilateral prefrontal lobotomy—S. D. Porteus and R. D. Kepner
2. A twin-controlled experiment on the learning of auxiliary languages—B. Price, W. J. Kostir, and W. M. Taylor

VOLUME 30—July-December, 1944

1. A method of administering and evaluating the thematic appreciation test in group situations—R. M. Clark
2. A study of anxiety reactions in young children by means of a projective technique—R. Temple and E. W. Amen

VOLUME 31—January-June, 1945

1. The evolution of intelligent behavior in rhesus monkeys—B. Weinstein
2. Perceptual behavior of brain-injured, mentally defective children: An experimental study by means of the Rorschach technique—H. Werner

VOLUME 32—July-December, 1945

1. A clinical study of sentiments: I.—H. A. Murray and C. D. Morgan
2. A clinical study of sentiments: II.—H. A. Murray and C. D. Morgan

VOLUME 33—January-June, 1946

1. Interpretation of spontaneous drawings and paintings—T. S. Waehner
Preferences for sex symbols and their personality correlates—K. Franck
2. Outstanding traits: In a selected college group, with some reference to career interests and war records—F. L. Wells and W. L. Woods

VOLUME 34—July-December, 1946

1. The relation of emotional adjustment to intellectual function—J. L. Despert and H. O. Pierce
The smiling response: A contribution to the ontogenesis of social relations—R. A. Spitz
2. Finger-painting and personality diagnosis—P. J. Napoli

VOLUME 35—January-June, 1947

1. The thematic apperception technique in the study of culture-personality relations—W. E. Henry
2. A continuation study of anxiety reactions in young children by means of a projective technique—M. Dorkey and E. W. Amen
A study of the vocational interest trends of secondary school and college women—A. M. Cawley

VOLUME 36—July-December, 1947

1. Maze test validation and psychosurgery—S. D. Porteus and H. N. Peters
2. The diagnostic implications of Rorschach's test in case studies of mental defectives—I. Jolles

VOLUME 37—January-June, 1948

1. The radio day time serial: A symbolic analysis—W. L. Warner and W. E. Henry
The relation of personality characteristics and response to verbal approval in a learning task—G. L. Grace
2. The mechanism of vision: XVIII. Effects of destroying the visual "associative areas" of the monkey—K. S. Lashley
A study of the relationship between handwriting and personality variables—P. Castelnuova-Tedesco

VOLUME 38—July-December, 1948

1. Modern language learning: The intensive course as sponsored by the United States Army, and implications for the undergraduate course of study—M. Lind
Conflict: A study of some interactions between appetite and aversion in the white rat—M. A. Tolcott
2. Schizophrenia and the *MAPS* test: A study of certain formal psycho-social aspects of fantasy production in schizophrenia as revealed by performance on the Make a Picture Story (*MAPS*) Test—E. S. Shneidman
A study of the transmission of authority patterns in the family—H. L. Ingersoll

VOLUME 39—January-June, 1949

1. A study of the psychoanalytic theory of psychosexual development—G. S. Blum
The assessment of parental attitudes in relation to child adjustment—E. J. Shoben, Jr.
2. Qualitative differences in the vocabulary responses of normals and abnormals—H. Feifel
The relative effectiveness of motion and still pictures as stimuli for eliciting fantasy stories about adolescent-parent relationships—P. E. Eiserer
The organization of hereditary maze-brightness and maze-dullness—L. V. Searle

VOLUME 40—July-December, 1949

1. An experimental study of what young school children expect from their teachers—B. Biber and C. Lewis
A study of the relative effects of age and of test difficulty upon factor patterns—H. A. Curtis
A projective experiment using incomplete stories with multiple choice endings—J. K. Seaton
2. Effects of sex role and social status on the early adolescent personality—E. Milner
Social perceptions and attitudes of children—M. Radke, H. Trager, and H. Davis

VOLUME 41—January-June, 1950

1. Some psychological and educational aspects of pediatric practice: A study of well-baby clinics—L. H. Blum
One-trial learning in the domestic rat—B. B. Hudson
An introduction to the principles of scientific psychoanalysis—A. Ellis
2. Awareness of racial differences by preschool children in Hawaii—D. V. Springer
Age trends in children's evaluation of teacher-approved and teacher-disapproved behavior—S. L. Witryol
The relationship between level of vocational aspiration and certain personal data: A study of some traits and influences bearing on the prestige level of vocational choice—J. Stubbins

VOLUME 42—July-December, 1950

1. Personality patterns of suicidal mental hospital patients—N. L. Farberow
Sex-rôle identification in young children in two diverse social groups—M. Rabban
2. A study of the influence of the social field on individual behavior: As revealed in the expression of hostility and warmth by neurotics and paranoid schizophrenics in discussion group situations—D. Shapiro
An experimental study of avoidance—R. F. Hefferline

VOLUME 43—January-June, 1951

1. A study of copying ability in children—E. A. Townsend
Prestige motivation of gifted children—D. P. Ausubel
2. A psychological study of physical scientists—A. Roe

VOLUME 44—July-December, 1951

1. The organization of hostility controls in various personality structures—S. Fisher and E. Hinds
Children and radio: A study of listeners and non-listeners to various types of radio programs in terms of selected ability, attitude, and behavior measures—E. A. Ricciuti
2. Quantitative expression in young children—W. E. Martin
The use of magnetic devices in the collection and analysis of the preverbal utterances of an infant—A. W. Lynip

VOLUME 45—January-June, 1952

1. Japanese-American personality and acculturation—W. Caudill
2. A statistical study of the Freudian theory of levels of psychosexual development—C. A. Barnes
Personality characteristics of selected disability groups—D. N. Wiener

Genetic Psychology Monographs *(continued)*

VOLUME 46—July-December, 1952

The relationship of social status, intelligence, and sex of ten- and eleven-year-old children to an awareness of poverty—F. J. Estvan
An empirical study of the castration and Oedipus complexes—S. M. Friedman
2. The relationship between projective test scoring categories and activity preferences—M. M. Schwartz
A comparison of formal and content factors in the diagnostic testing of schizophrenia—M. Sherman

VOLUME 47—January-June, 1953

1. Ability and accomplishment of persons earlier judged mentally deficient—D. C. Charles
Variations in the consistency of the behavioral meaning of personality test scores—M. Kornreich
2. Some child-rearing antecedents of aggression and dependency in young children—R. R. Sears, *et al.*
Symptom correlates for descriptive diagnosis—J. R. Wittenborn, *et al.*

VOLUME 48—July-December, 1953

1. Age and mental abilities: A longitudinal study—W. A. Owens, Jr.
The development of a personality questionnaire for drinkers—P. J. Hampton
Personality and physical disease: A test of the Dunbar hypothesis applied to diabetes mellitus and rheumatic fever—D. H. Crowell
2. Socio-economic contrasts in children's peer culture prestige values—B. Pope
A critical review of the stability of social acceptability scores obtained with the partial-rank-order and the paired-comparison scales—S. A. Witryol and G. G. Thompson
A study of the effects of color on Rorschach responses—G. G. Brody

VOLUME 49—January-June, 1954

1. Factors underlying major reading disabilities at the college level—J. A. Holmes
Parent behavior toward first and second children—J. K. Lasko
2. Social-status and intelligence: An experimental study of certain cultural determinants of measured intelligence—E. A. Haggard
Certain determinants and correlates of authoritarianism—S. Siegel
Personalities in faces: I. An experiment in social perceiving—P. F. Secord, W. F. Dukes, and W. Bevan

VOLUME 50—July-December, 1954

1. A study of the relationship between play patterns and anxiety in young children—E. W. Amen and N. Renison
Operational exploration of the conceptual self system and of the interaction between frames of reference—M. Edelson and A. E. Jones
Problem solving: A statistical description of some relationships between organismic factors and selected response measures—N. A. Fattu, E. Kapos, and E. V. Mech
2. The relation of cortical potentials to perceptual functions—C. Chyatte
The import for clinical psychology of the use of tests derived from theories about infantile sexuality and adult character—D. W. Miles
Measuring personality in developmental terms: The Personal Preference Scale.—M. H. Krout and J. K. Tabin

VOLUME 51—January-June, 1955

1. Some relations between techniques of feeding and training during infancy and certain behavior in childhood—A. Bernstein
The expression of personality in drawings and paintings—L. H. Stewart
2. Negative stereotypes concerning Americans among American-born children receiving various types of minority-group education—J. A. Fishman
The Lincoln-Oseretsky Motor Development Scale—W. Sloan

VOLUME 52—July-December, 1955

1. Some personality correlates of sex, sibling position, and sex of sibling among five- and six-year-old children—H. L. Koch
A quantitative Rorschach assessment of maladjustment and rigidity in acculturating Japanese Americans—G. DeVos
Measurement of authoritarianism and its relation to teachers' classroom behavior—H. M. McGee
2. The formal aspects of schizophrenic verbal communication—B. Mirin
A study in an aspect of concept formation, with subnormal, average, and superior adolescents—H. N. Hoffman
Traumatic avoidance learning: Acquisition and extinction in dogs deprived of normal peripheral autonomic function—L. C. Wynne and R. L. Solomon

VOLUME 53—January-June, 1956

1. As the psychiatric aide sees his work and problems—F. L. Wells, M. Greenblatt, and R. W. Hyde
An investigation of avoidance, anxiety, and escape behavior in human subjects as measured by action potentials in muscle—J. D. Brothers
Spread of effect: A critical review—M. H. Marx
2. Stress, fantasy, and schizophrenia: A study of the adaptive processes—O. J. B. Kerner
The attitude structure of the individual: A Q-study of the educational attitudes of professors and laymen—F. N. Kerlinger

VOLUME 54—July-December 1956

1. A study of personality differences between middle and lower class adolescents: The Szondi Test in culture-personality research—L. Rainwater
The assessment of parental identification—S. W. Gray and R. Klaus
2. The influence of social context on impulse and control tendencies in preadolescents—G. H. Zuk
Tender-mindedness versus tough-mindedness in psychology: A reexamination—H. Winthrop
A method for the comparison of groups: A study in thematic apperception—L. C. Schaw and W. E. Henry

VOLUME 55—January-June, 1957

1. Academic performance and personality adjustments of highly intelligent college students—B. M. Horrall
The use of the Vineland Social Maturity Scale in the planning of an educational program for non-institutionalized low-grade mentally deficient children—M. G. Gottsegen
2. The structure and origin of the anal character—H. Beloff
Free expression of adolescents' interests—M. Amatora
The rôle of mass media and the effect of aggressive film content upon children's aggressive responses and identification choices—R. S. Albert
Interest in persons as an aspect of sex difference in the early years—E. W. Goodenough

VOLUME 56—July-December, 1957

1. Some stable response determinants of perception, thinking, and learning: A study based on the analysis of a single test—M. L. Simmel and S. Counts
Potentials of age: An exploratory field study—S. L. Pressey
2. Tactile communication—L. K. Frank
The use of a filmed puppet show as a group projective technique for children—M. R. Haworth
The social competence of middle-aged people—R. J. Havighurst

VOLUME 57—January-June, 1958

1. Psychological and cultural problems in mental subnormality: A review of research—S. B. Sarason and T. Gladwin
2. Developmental aspects of discrimination in relation to adjustment—P. London
Muscular tension as a response to psychological stress in rheumatoid arthritis and peptic ulcer—J. A. Southworth
Frustration: Theory and experiment—R. Lawson and M. H. Marx

VOLUME 58—July-December, 1958

1. Emotional aspects of political behavior: The woman voter—E. M. Bennett and H. M. Goodwin
The accuracy of self estimate: A measure of self-concept reality—R. M. Brandt
Personality factors in social mobility: A study of occupationally mobile businessmen—J. C. Abegglen
2. Interaction between authoritarian and nonauthoritarian principals and teachers—P. Lambert
Some factors related to the choice-status of ninety eighth-grade children in a school society—D. Elkins

Genetic Psychology Monographs *(continued)*

VOLUME 59—January-June, 1959

1. Freshman Rorschachs and college performance—G. R. Schmeidler, M. J. Nelson, and M. Bristol
Perceptual aspects of attributed movement—R. W. Coan
Men and women: Personality patterns and contrasts—E. M. Bennett and L. R. Cohen
2. Personality factors in mothers of cerebral palsied children—G. Boles
Separatism and integrationism: A social-psychological analysis of editorial content in New York newspapers of three American minority groups—J. A. Fishman and G. S. Fishman
Self, role, and satisfaction—A. L. Brophy

VOLUME 60—July-December, 1959

1. The constancy of personality ratings over two decades—R. D. Tuddenham
Mother-son identification and vocational interest—L. H. Stewart
A study of mother-child relationships in the emotional disorders of children—M. J. Rosenthal, M. Finkelstein, E. Ni, and R. E. Robertson
On the trail of the wolf-children—W. F. Ogburn and N. K. Bose
2. Measuring the mental health of normal adults—R. F. Peck
Parental identification in young children—W. Emmerich
The problems of Jewish religious youth—B. M. Levinson

VOLUME 61—January-June, 1960

1. Identifying socially maladjusted school children—A. A. Adams
Social interaction in an interracial nursery school—H. W. Stevenson and N. G. Stevenson, *et al.*
Interests of pre-adolescent boys and girls—M. Amatora
Educational and sex differences in the organization of abilities in technical and academic students in Colombia, South America—J. F. Filella
2. Some themes in the personalities of German men—L. Rainwater
Right-left gradients in body image, body reactivity, and perception—S. Fisher
Longitudinal survey of child Rorschach responses: Younger subjects two to 10 years—L. B. Ames
Personality patterns related to occupational roles—M. Siegelman and R. F. Peck

VOLUME 62—July-December, 1960

1. The reading process: A descriptive, interdisciplinary approach—M. P. Langman
Reactions to stress as a function of the level of intelligence—S. M. H. Zaidi
Rorschach diagnosis by a systematic combining of content, thought process, and determinant scales—P. A. Bower, R. Testin, and A. Roberts
Longitudinal survey of child Rorschach responses: Older subjects aged 10 to 16 years—L. B. Ames
2. Twenty years of shock therapy in America, 1937-1956: An annotated bibliography—N. H. Pronko, R. Sitterly, and K. Berg
The quantitative analysis of parent behavior toward psychotic children and their siblings—E. M. Donnelly

VOLUME 63—January-June, 1961

1. Make a sentence test: An approach to objective scoring of sentence completions—E. F. Borgatta
Some aspects of the relationship between perception and motility in children—E. P. Rothman
Some psychophysical aspects of mental illness in children—P. H. Berkowitz
Conceptual thinking in schizophrenic children—G. Friedman
2. Some maternal influences on children's personality and character—J. C. Finney
A study of concept formation in normals, mental defectives, and brain-damaged adults—R. W. Zaslow
An investigation of trait relationships among six-year-old children—S. A. Avakian

VOLUME 64—July-December, 1961

1. Coping patterns of preschool children in response to intelligence test demands—A. Moriarty
Studies of individual and paired interactional problem-solving behavior of rats: II. Solitary and social controls—A. Baron and R. A. Littman
2. Studies of role performance—J. H. Mann
A psychological assessment of professional actors and related professions—R. Taft
A cross-cultural study of menstrual taboos—W. N. Stephens
The biological roots of creativity—H. Gutman

VOLUME 65—January-June, 1962

1. A Rorschach study of the development of personality structure in white and Negro children in a Southeastern community—A. C. Price
Measuring personality patterns of women—J. Loevinger
Home interests in early adolescence—M. Amatora
Social and emotional adjustment during adolescence as related to the development of psychosomatic illness in adulthood—L. H. Stewart
2. A systematic study of interaction process scores, peer and self-assessments, personality and other variables—E. F. Borgatta
Children's conceptions of the content and functions of the human body—E. Gellert

VOLUME 66—July-December, 1962

1. Ocular-manual laterality and perceptual rotation of literal symbols—I. Flescher
Parental child-rearing attitudes and their correlates in adolescent hostility—S. B. Chorost
Some explorations into the nature of anxieties relating to illness—P. R. Robbins
Developmental-interactive aspects of hypnosis: Some illustrative cases—J. R. Hilgard and E. R. Hilgard
2. Teaching a two-year-old to read: An experiment in early childhood learning—W. Fowler
Achievement gains or losses during the academic year and over the summer vacation period: A study of trends in achievement by sex and grade level among students of average intelligence—K. M. Parsley, Jr., and M. Powell

VOLUME 67—January-June, 1963

1. Verbatim dialogue of a mother and child in therapy—C. E. Moustakas
A study of self-concept: Delinquent boys' accuracy in selected self-evaluations—W. E. Amos
An integrating study of the factor structure of adult attitude-interests—R. B. Cattell and J. Horn
Social factors in mental disorders: A contribution to the etiology of mental illness—W. G. Cobliner
2. Conceptual preferences associated with choice of quantitative and nonquantitative college majors—J. Gilbert
A study of the concept of the feminine role of 51 middle-class American families—A. Steinmann

VOLUME 68—July-December, 1963

1. Brain dynamism as reflected in illusions—G. K. Yacorzynski
The nature and measurement of components of motivation—R. B. Cattell, J. A. Radcliffe, and A. B. Sweney
2. Sharing in preschool children as a function of amount and type of reinforcement—W. F. Fischer
The Gesell Incomplete Man Test as a measure of developmental status—L. B. Ames and F. L. Ilg
Psychological and sociological factors in prediction of career patterns of women—M. C. Mulvey
Functionalism and its historical significance—R. Harrison

PUBLICATION IN THE JOURNAL OF PSYCHOLOGY

The Journal of Psychology provides a unique service to authors. Articles accepted for publication in this periodical appear in print with the least possible amount of delay. In line with this policy, all steps of the acceptance procedure are maximally streamlined, and the author's work is presented with a bare minimum of editing or restriction of any kind. *The Journal of Psychology* is the ideal medium for presentation of timely professional material.

Authors are requested to exercise unusual care in the preparation of papers for this journal. Accuracy, grammar, and style should be checked and rechecked before a clean version of the manuscript is submitted. Careful attention should be paid to the General Instructions on the back cover of each issue and to the announcement on the inside front cover.

Authors are particularly requested to note the fact that all parts of the manuscript—including references, footnotes, and tables—must be double-spaced.

$30.00 per annum
$15.00 per volume
$7.50 per half volume

BIMONTHLY
Three volumes per year
Immediate publication

May, 1969
Volume 72, First Half

(Founded in 1935 by Carl Murchison)

The Journal of PSYCHOLOGY

The General Field of Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from the date of publication.

MAY, 1969

(Manuscripts are printed in the order of final acceptance)

Quantitative analysis of psychoanalytic studies of Shakespeare 3
By M. S. Lindauer

T-group member ratings of leader and self in a human relations laboratory . 11
By Earle A. Koile and Carolyn Draeger

The relationship of illness and health behavior to paper-and-pencil tests . . 21
By Jean S. Kerrick and Virginia A. Clark

A note on the use of discrepancy scores in a self-concept inventory . . . 33
By Shirin Schludermann and Eduard Schludermann

Prediction of supervisory success by use of standard psychological tests . . 35
By H. Edwin Titus

Reversal and nonreversal shifts in neurologically handicapped children . . 41
By Robert H. Reiher, Joseph G. Phelan, and Vernon L. Kiker, Jr.

(over)

Copyright © 1969 by The Journal Press
Provincetown, Massachusetts, U. S. A.
All rights reserved

Published at 35 New St., Worcester, Mass.
Second-class postage paid at Worcester, Mass.
Send all mail to Managing Editor, 2 Commercial Street, Provincetown, Massachusetts 02657

The study of self-disclosure in experimental interviews 55
By Fred W. Vondracek

An analysis of the instructional effectiveness of visual illustrations presented via television 61
By Francis M. Dwyer, Jr.

Field independence and performance in distraction 65
By B. Kent Houston

Contrasting personality patterns of high and low achievers among college students of Italian and Irish descent 71
By Angelo Danesino and William A. Layman

Explorations in methodology: The scientific question as a schemapiric dimension 85
By Sven Lundstedt

A comparison of techniques for the scoring of avoidable errors on the Raven Progressive Matrices 93
By John F. Fracchia, Diane Fiorentino, Charles Sheppard, and Sydney Merlis

Laymen's identification of item content on the Eysenck Personality Inventory . 99
By Gordon Stanley

The factorial determination of types of conforming individuals . . . 101
By Saundra F. Zimmerman, Darhl M. Pedersen, and Kay Smith

Biographical inventories and style of leadership 109
By John J. La Gaipa

Untimed norms for the SRA Mechanical Knowledge and Shop Arithmetic Tests 115
By Kentner V. Fritz and Martin F. Fritz

The role of verbal factor in a forced-choice test of achievement value . . 119
By Bishwa Nath Mukherjee

An investigation of success predictors in graduate school in psychology . . 125
By James R. Rawls, Donna J. Rawls, and C. Wade Harrison

The effect of sample items on test performance 131
By Donald W. Felker, Gerald Larson, Frank Milhollan, and John R. Pyper

Comparison between American permanent residents of Israel: Part II, Israeli background 135
By Gerald Engel

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 3-9.

QUANTITATIVE ANALYSES OF PSYCHOANALYTIC STUDIES OF SHAKESPEARE*

Department of Psychology, State University of New York, Brockport

---

M. S. LINDAUER

"*Whenever you can, count*"—Francis Galton.
"*You may prove anything by figures*"—Thomas Carlyle.
"*I am ill at reckoning*"—Shakespeare.

A. INTRODUCTION

Holland's extensive review and analysis of psychoanalytic studies of Shakespeare (4) convincingly argues for the contribution of psychoanalysis to literature. "The insights of psychoanalysis complement and fulfill those of conventional criticism" (4, p. 347). Yet Holland's effort, sophisticated and articulate as it is, is more of an artistic and literary performance than a scientific and objective accomplishment. As a man of literature and not of science, Holland's approach is justified and perhaps appropriate in handling the quasi-scientific/artistic status of psychoanalysis. "Psychoanalysis represents a most important strategic bridge from the rigorous, external aims of the sciences, and those internal aims of the humanities, less rigorous but more significant and, if you will, more practical" (4, p. 348).

The absence of the scientific method has been noted by Plank in his review of Holland's book (15) as the serious flaw in psychoanalytic studies of Shakespeare. Holland's treatment of the role of psychoanalysis shares this shortcoming. In particular, it is the absence of quantitative analysis, the scientist's tool of exactness and precision, which limits the information that could be abstracted from masses of raw data. Statistical techniques systematically describe and relate data, and allow inductive statements regarding the probability of events to be made: i.e., whether observations are due merely to chance or to significant factors. Questions can thereby be answered with some assurance, and promising new inquiries can be initiated. In a sense, such a strategy follows Shakespeare's own dictum: "Be sure of it; give me ocular proof" (Othello).

Relatively few academic psychologists (in contrast to psychoanalysts and

---

[*] Received in the Editorial Office, Provincetown, Massachusetts, on January 9, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

clinical psychologists) have sympathetically seen an interrelationship between psychology and literature (1, 2, 8). Most psychologists have viewed the relationship between the two subjects as one of incompatibility, alienation, and failure (6, 17, 20), mainly because of literature's unsusceptibility to standard technical analyses. Nevertheless, some attempts have been made to apply quantitative methods to the study of literature (9, 10, 11, 18), including Shakespeare (12, 19). Yet even one favorably disposed to such an analysis has been less than enthusiastic over the results (15, p. 202).

The quantitative approach to qualitative material evokes the controversy between the statistical *versus* the clinical approach to man. Miller (14) has posed the disparity between these two strategies in this manner:

"For several years there has been a running battle between the clinical heirs of Sigmund Freud and the statistical heirs of Sir Francis Galton. The Freudians learn about people by talking to them; the Galtonians give tests and compute statistics. When both groups are not busy doing this, they like to spend their time criticizing each other" (14, p. 315).

Miller further illustrates this dilemma as follows (14, p. 317): clinicians attribute to their method such terms as "dynamic, rich, genuine," while they attach to the statistical method the labels "mechanical, artificial, trivial." On the other hand, statisticians call themselves "verifiable, reliable, testable" and claim the clinicians are "mystical, primitive, muddleheaded."

The purpose of this paper is to investigate an aspect of psychoanalysis, as found within the context of its analysis of Shakespeare, not pursued by Holland. He focused on studies of Shakespeare the man, his works, the audience, and the creative process. But Holland missed the opportunity to disclose something of psychoanalysis, as its interests are revealed in the study of Shakespeare. The present inquiry, using quantitative techniques, is addressed to this issue.

B. METHOD

The subject index and listings of books and articles in Kiell's exhaustive bibliography of psychological studies of literature (7) were examined for references to Shakespearean plays and characters, as well as psychological traits dealt with in the plays and characters. The index was considered a reliable indicator of the main emphasis and concern of a study, since the citation was based not only on the title of the work, but on the editor's reading of either the reference itself, its abstract, or a review (7, p. 5). Although Shakespearean material may have been mentioned in psychological works other than those included in this survey, omission from the index of a study would indicate that treatment of Shakespeare in the study was peripheral to some other topic or issue.[1] The psychoanalytic status of a work, generally defined so as to include all varieties and schools of depth analysis, was determined either through the index reference or by a reading of the title. While a few nonpsychoanalytical works may therefore have been accidentally included, over 98 per cent of the studies could be rather easily categorized as psychoanalytic. In addition, early (pre-Freud) writings were also included, so as to serve as a base for later comparisons in the trends-over-time analysis.

C. RESULTS

1. *Preferences for Plays*

Shakespeare wrote 37 plays (16), 65 per cent or 24 of which have been specifically studied in detail by psychoanalysts in 196 publications. With 84 additional studies concentrating on Shakespeare himself, psychoanalysis appears disproportionately more interested in the man than in any one category of his works ($\chi^2 = 4.00$, $df = 1$, $p < .05 > .01$). Arranging the plays of interest to psychoanalysis into the categories Comedy, Tragedy, and History (5), the 24 plays are distributed as follows: 9, 11, and 4, respectively (out of a total possible breakdown of 14, 13, and 10, respectively).[2] Although it would appear that historical and comic categories were of less interest than tragedies, the differences are not statistically significant ($\chi^2 = 0.06$, $df = 1$, $p > .05$). Nor is there any statistical difference, in terms of these three categories, between the number of plays written by Shakespeare and the number studied by psychoanalysts ($\chi^2 = 1.76$, $df = 1$, $p > .05$). However, with respect to the number of studies devoted to plays in these three categories, more effort was spent on the tragedies (79 per cent of the 196 studies) than on the comedies (15 per cent) or histories (6 per cent); these differences are statistically significant ($\chi^2 = 164.67$, $df = 1$, $p < .01$). Of the tragedies, *Hamlet* is easily the most preferred, accounting for 58 per cent. Together with *Macbeth, Othello* and *King Lear,* these four tragedies account for 72 per cent of all studies.

2. *Time Trends*

Fifty per cent of all psychoanalytic studies of Shakespeare and his works have occurred since 1943. This interest is increasing: 3 per cent, 9 per cent, 16 per cent, 16 per cent, 23 per cent and 27 per cent of all studies, respectively, were published in six 10-year intervals beginning in 1903. This overall trend is also reflected in the subtotals that refer to the three categories of plays, as well as those studies concentrating on Shakespeare. Comedies and histories have shown the greatest spurt in the most immediate decade (1953-1962): 31 per cent of the former type and 54 per cent of the latter were published then. Studies of tragedies and of Shakespeare have shown relative stability over the past two-three decades. Over the entire time span, studies of tragedy have outnumbered those of Shakespeare by at least if not more than 2:1 (in 1953-1962, the percentages were 55:25) except for 1933-1942 (42:40) and a period before Freud, 1851-1902 (38:62). Finally, psychoanalysts have not shown any special interest in plays written in any one particular period of Shakespeare's career. Thus, plays written during Shakespeare's early, middle, and late years (16), received about equal amounts of attention, as shown by the lack of a significant correlation between date of authorship and frequency of analytic studies ($r_{\text{rho}} = -.13$, $p > .05$).

---

[1] This helps to account for disagreements between Kiell (7) and Holland's (4) coverage, since the latter apparently included psychological material which made any reference to Shakespeare.

[2] The six omitted historical plays were all in the *Henry* series.

3. *The Characters Studied*

There are 325 characters in Shakespeare's plays (13).[3] This is an average (mean) of 8.78 characters per play: historical plays have the most (10.80) and tragic plays the fewest characters (7.61). In 33 studies, psychoanalysts have specifically focused on the characters in 12 plays, concentrating on only about 6 per cent of the total possible. Interest has been particularly directed toward the tragic figures ($\chi^2 = 10.12$, $df = 1$, $p < .01$), accounting for 73 per cent, 50 per cent, and 63 per cent of the above studies, plays, and characters, respectively.

4. *The Traits*

In 136 studies, psychoanalytic authors specifically emphasized 60 psychological traits, about half or 28 of which referred to "psychopathological states" (e.g., *insanity*), with the remainder about equally divided between the categories "symbols" (*myths*), "processes" (*the unconscious*), "motives" (*love*), and "developmental" (*psychosexual stages*). The most frequently referred to traits (in six-11 studies), in decreasing order, were *madness, homosexuality, Oedipus complex, dreams,* and *fantasy*. The plays arranged by category, and the studies of Shakespeare, showed no unexpected differences in the distribution of these traits, except for "psychopathology" which was referred to more often in tragedies and less often in Shakespeare than expected ($\chi^2 = 15.08$, $df = 3$, $p < .01$).

---

[3] Titled characters—e.g., "courtesan"—were excluded from the count.

5. *Freud and Other Analysts*

In his writings Freud specifically referred to about half (18) of Shakespeare's plays (4, pp. 359-365). In terms of the three categories into which Shakespeare's plays can be placed, no one type predominated in Freud's writings, nor do the categories of plays studied by Freud differ from those not studied by him ($\chi^2 = 1.67$, $df = 2$, $p > .05$). Compared to plays studied by psychoanalysts in general, Freud made no mention of 14 plays noted by other analysts. Particularly noteworthy, in that six studies were devoted to each, are *Coriolanus, All's Well That Ends Well,* and *The Winter's Tale;* Freud referred to one play not studied by other analysts (*Love's Labour's Lost*). Both Freud and the analysts jointly referred to 18 plays (4, pp. 371-390), showing remarkable agreement in the attention given to them: the difference between the two in the number of studies associated with each play (arranged into ranks) was not significant ($t = 0.18$, $df = 17$, $p > .05$); *e.g., Hamlet* and *Macbeth* received the most attention from both Freud and the analysts, *Henry 6* and *Much Ado About Nothing* received little, and *Julius Caesar, Richard 3,* and *Romeo and Juliet* received intermediate amounts of interest.

D. DISCUSSION AND CONCLUSIONS

The following picture of psychoanalysis emerges from a quantitative examination of its treatment of Shakespeare and his works. More interested in the man than in any one category of his works, psychoanalysts have studied a representative proportion but not nearly all of Shakespeare's plays. Of those studied, while psychoanalysis makes no distinctions between works produced at various periods in Shakespeare's life, or between the comic, tragic, and historical categories of plays, not all plays have received the same consideration: tragedies are preferred. Of the tragedies studied, not all have received the same degree of attention. This somewhat narrow and confined approach to the plays is also seen in psychoanalysis' concentration on a small number of characters in the plays, with tragic figures receiving the most emphasis. Psychoanalytic interests, which are with some notable exceptions paralleled by Freud himself, are displayed more frequently in relatively recent years, with some categories of plays showing a leveling of interest, while others display a sharp increase. Tragic plays, with a few exceptions, have consistently been of the greatest interest over the years. Finally, psychoanalysts have concentrated more on the morbid rather than the "normal" aspects of Freud's contribution to the study of man, this being particularly characteristic of the tragedies.

These aspects of psychoanalysis, derived from quantitative analyses, raise certain questions that are not within the province of this study to answer, but which presumably are more amenable to literary rather than quantitative attack. Firstly, what reasons can be ascribed to the relative if not absolute neglect by psychoanalysis of many of Shakespeare's plays, as well as characters? Related to this question is the glaring disparity between those few plays solely of interest to psychoanalysts at large but not, evidently, to Freud. Are some plays of less literary merit; or do they lack psychoanalytic implications? Secondly, the great interest evidenced in the tragedies may not be surprising, but why the neglect of the comedies? Freud has indicated the importance of humor in understanding the deep-seated forces in man (3). For that matter, Freud's interest in the role of the past in shaping the present and the future is perhaps contradicted by the slighting by psychoanalysis of the historical plays. In the same sense, the lack of any special concern with Shakespeare's early writings seems surprising. More of the man and his antecedents might be revealed more clearly in the early rather than later productions. Finally, psychoanalysis is a theory not just of neurosis but of the normal personality as well. Yet psychoanalysts—as is perhaps true of most authors, as well as the man-in-the-street—have ignored the broader aspects of Freud in order to focus on the abnormal.

It is hoped that this quantitative approach may be evaluated in light of the questions answered and those raised—questions that might not have been answered or could not have been raised by a qualitative approach. However, the ease with which figures may be statistically manipulated is a danger, especially if data are forced into convenient molds for purposes of measurement rather than for their usefulness and clarity.

> There is no more common error than to assume that, because prolonged and accurate mathematical measurements have been made, the application of the result to some fact of nature is absolutely certain—A. N. Whitehead.

E. Summary

The application of statistical techniques to psychoanalytic studies of Shakespeare and his works reveals the preferences of psychoanalysis for plays, characters, and traits; trends over time; and similarities and distinctions between Freud and other analysts. Questions about some of the preoccupations and omissions of psychoanalysis in this area are raised on the basis of the analysis, as well as the role of quantification in the psychological study of literature.

References

1. Allport, G. W. Personality. New York: Holt, 1937.
2. Cohen, J. Humanistic Psychology. New York: Collier, 1962.

3. Freud, S. Wit and its relation to the unconscious. In *The Basic Writings of Sigmund Freud* (trans., A. A. Brill). New York: Modern Library, 1938.
4. Holland, N. N. Psychoanalysis and Shakespeare. New York: McGraw-Hill, 1966.
5. Irvin, T. A Pronouncing Dictionary of Shakespearean Proper Names. New York: Barnes & Noble, 1945.
6. Johnson, D. M. Psychology versus literature. *Harper Books & Authors,* 1961, **12**, 1-14.
7. Kiell, N., *Ed.* Psychoanalysis, Psychiatry and Literature: A Bibliography. Madison: Univ. Wisconsin, 1963.
8. Levitas, G. B., *Ed.* The World of Psychology, Vol. I & II. New York: Braziller, 1963.
9. Lindauer, M. S. The nature and use of the cliché. *J. Gen. Psychol.,* 1968, **78**, 133-143.
10. McCurdy, H. G. A study of the novels of Charlotte and Emily Bronte as an expression of their personalities. *J. Personal.,* 1947, **26**, 109-152.
11. ———. A mathematical aspect of fictional literature pertinent to McDougall's theory of a hierarchy of sentiments. *J. Personal.,* 1948, **17**, 75-82.
12. ———. The Personality of Shakespeare: A Venture in Psychological Method. New Haven: Yale, 1953.
13. Magill, F. N., *Ed.* Masterpieces of World Literature in Digest Form. New York: Harper & Row, 1952 (1st series); 1955 (2nd series).
14. Miller, G. A. Psychology: The Science of Mental Life. New York: Harper & Row, 1962.
15. Plank, R. Review of Holland's *Psychoanalysis and Shakespeare. Lit. & Psychol.,* 1966, **16**(3 & 4), 200-203.
16. Reese, M. M. Shakespeare: His World and His Work. New York: St. Martin's, 1953.
17. Royce, J. R. Metaphoric knowledge and humanistic psychology. In J. F. T. Bugental (Ed.), *Challenges of Humanistic Psychology.* New York: McGraw-Hill, 1967.
18. Skinner, B. F. A quantitative estimate of certain types of sound patterning in poetry. *Amer. J. Psychol.,* 1941, **54**, 64-79.
19. ———. The alliteration in Shakespeare's Sonnets: A study in literary behavior. *Psychol. Rec.,* 1939, **3**, 186-192.
20. Wyatt, F. Psychology and the humanities: A case of no-relationship. In J. F. T. Bugental (Ed.), *Challenges of Humanistic Psychology.* New York: McGraw-Hill, 1967.

Department of Psychology
State University College at Brockport
Brockport, New York 14420

Published as a separate and in *The Journal of Psychology,* 1969, **72,** 11-20.

T-GROUP MEMBER RATINGS OF LEADER AND SELF IN A HUMAN RELATIONS LABORATORY*[1]

Department of Educational Psychology, The University of Texas; and Department of Counseling, The University of Maryland

EARL A. KOILE AND CAROLYN DRAEGER

A. PURPOSES

Changes in human behavior continue to be extremely difficult to measure and yet intriguing to discover, whether through carefully designed research or by simple observation. The present study focuses on change. It is an investigation of how members of a human relations laboratory change in their perceptions of their group leaders during different T-Group (training group) sessions and in their perceptions of themselves as the laboratory experience progresses. More specifically the study investigates whether or not (*a*) members of T Groups and their group leaders become more like each other during the laboratory, (*b*) members see their leaders differently from session to session, (*c*) members become more like each other in the ways they rate their leaders as the T Groups continue to meet, (*d*) member ratings of T-Group leaders become more like the leader's ratings of himself as the sessions continue, and (*e*) whether or not members tend to rate themselves differently at the end as compared to the beginning of the human relations laboratory experience.

B. METHOD

The sample for the study consisted of participants in a four-day human relations laboratory held for undergraduates at the University of Texas at Austin.[2] The 44 group members met together for eight general sessions conducted by the laboratory director, and were divided into four T Groups for six sessions with four experienced T-Group leaders.

The laboratory general sessions and T Groups were experientially oriented toward the personal growth of members. Theoretical implications and gen-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 31, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Appreciation is expressed to Shirley Bird Perry and Jack Steele of The University of Texas Student Union for their contributions to this study.

[2] Data used in this study were analyzed and reported differently in a thesis (1) supervised by the senior author.

eralizations tended to follow rather than to precede experiences. Activities during general sessions included brief discussions in small encounter groups, role playing, psychodramatic scenes, demonstration groups, and a variety of other verbal and nonverbal encounters. Thus the laboratory experience moved from general sessions to T Groups and back to general sessions to foster the personal development of individual members and to create a climate in which participants, hopefully, could discover freedom to explore relationships with others, their feelings about themselves, and patterns of their coping behavior, and could learn to give, receive, and use feedback constructively.

The rationale of the Semantic Differential (5) was used in developing an instrument consisting of 30 bipolar adjective pairs. The two adjectives in each pair were opposite poles of a seven-point rating scale. Each group member and leader used the 30 adjective scales (*a*) to rate the T-Group leader at the conclusion of each of the six T-Group sessions and (*b*) to rate himself in the beginning and in the concluding general session of the laboratory.

In scoring the instrument, a score of +7 represented a rating at the positive end of each adjective scale (e.g., warm); a score of +1 represented a rating at the negative end (e.g., cold). Ratings on the 30 scales were considered separately, as well as in sum, in the statistical design of the study. The analysis of variance method with single classification and repeated measurements was used to test for statistical significance, except where similarity among members was of interest and a test for the significance of the difference between correlated variances was appropriate.

C. RESULTS AND DISCUSSION

1. *Overall, T-Group members and leaders did not rate themselves more similarly at the end compared to the beginning of the group experience.* Mean absolute differences between self-ratings of each member and his leader at the beginning and again at the end of the laboratory were computed, and the overall means were compared by analysis of variance. The difference between the initial and the final mean sum of member-leader differences on the 30 adjective scales was not significant. (Initial and final mean sums of member-leader differences were 43.36 and 44.20 respectively.)

The finding that T-Group members and leaders did not become more alike suggests that members may not have been modeling themselves after the group leader, even though this is often regarded as membership tendency and as a correlate of change (6). Or, members may have already been sufficiently like their leader and, therefore, were unlikely to become more like him. It is conceivable that influences of the general sessions combined with the climate in the T Groups and in the laboratory overall may have reduced the need for member-leader modeling.

2. *Laboratory members changed their ratings of the T-Group leaders from session to session, thus suggesting that member perceptions and, by inference, the group climate were continuously changing.* An analysis of variance comparison of mean sums of scores representing members' ratings of the group leaders in the six T-Group sessions yielded an F ratio of 14.10 which, with 5 and 215 degrees of freedom, was significant far beyond the .001 level. Moreover, in the comparison of mean ratings on each of the 30 adjective scales, the change in members' ratings of the group leaders over the six sessions was found to be significant beyond the .05 level on 25 of the 30 scales.

a. Laboratory members saw the T-Group leaders in a more positive manner in the last T-Group meeting than they did in the first meeting. An analysis of variance comparison indicated a highly significant difference between the mean total scores representing members' ratings of the group leader in the first and last meetings. (First and last session means were 143.36 and 162.18 respectively; $F = 23.47$; $df = 1{,}43$; $p < .001$.) Moreover, there were significant differences between the first and last session ratings on 19 of the 30 adjective scales. Qualitatively speaking in the last, compared to the first T-Group meeting, members regarded the group leaders as warmer, deeper, and more open, reaching out, trusting, involved, sensitive, supporting, articulate, understanding, caring, insightful, accepting, natural, authentic, forthright, spontaneous, sharing, and impulsive.[3]

The finding that members came to perceive the group leaders more positively supports a similar finding in another study (3) conducted under the senior author's supervision, in which the mean member rating of the group leader was also found to be significantly more positive in the last compared to the first T-Group session. Moreover, members continued to regard the group leader in a significantly more positive manner six weeks after the conclusion of the laboratory than they had at its outset.

b. While there was an overall positive shift in the mean total scores representing members' ratings of the group leaders from the first to the last session, this shift was not continuously positive. Total score means for the six T-Group sessions are plotted in Figure 1 and analysis of variance data on significant comparisons are provided.

In the second session, as may be noted from Figure 1, T-Group members egarded the leaders in a significantly more negative manner than in the pre- eding session. There was also a slightly negative mean shift from the fourth o fifth sessions, but statistical significance was not attained. The positive shifts n mean member ratings of the group leaders from the second to third, third to fourth, and fifth to sixth T-Group sessions were all statistically significant.

---

[3] The mean shift on the scale deliberate-impulsive was the only significant shift in a negative direction. It should be noted that the final mean, 4.80, was still toward the "positive" adjective deliberate.

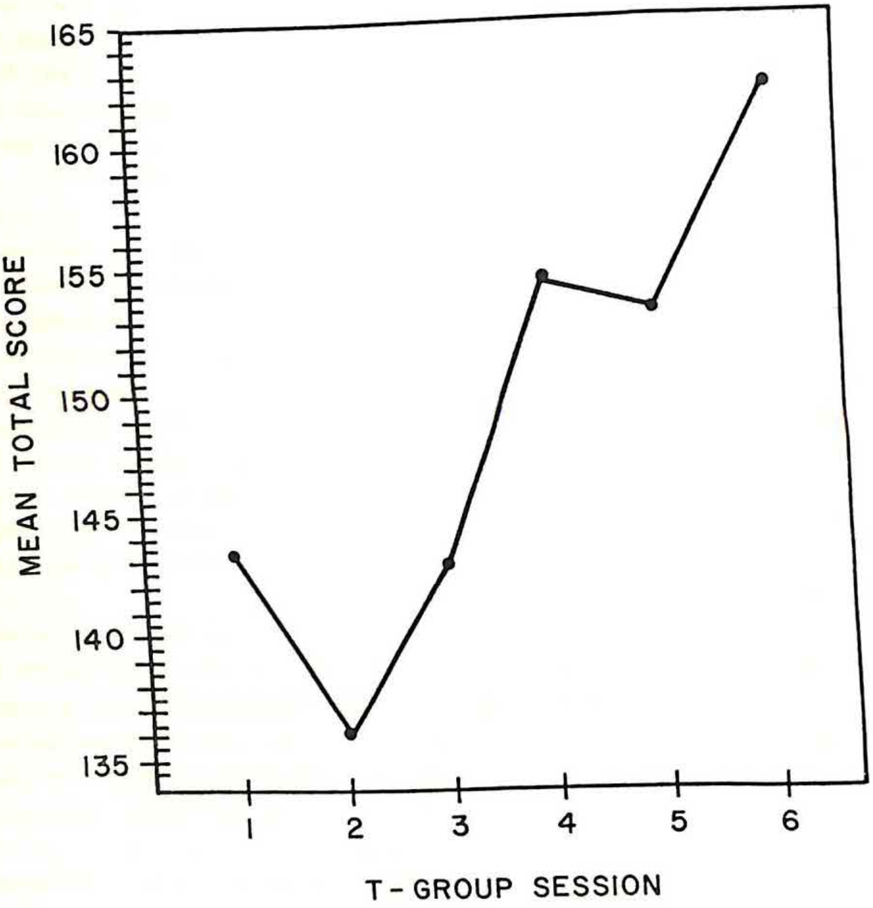


FIGURE 1

Plot of Mean Total Scores Representing Members' Ratings of the Group Leader for Each T-Group Session

Analysis of variance data on the significant comparisons between means are given below in the following order: sessions compared; respective means; F ratio; and p with $df = 1,43$.

1 and 2, 143.36 and 136.02, 4.01, $< .05$.
2 and 3, 136.02 and 142.98, 5.69, $< .05$.
3 and 4, 142.98 and 154.25, 17.84, $< .001$.
5 and 6, 153.14 and 162.18, 6.04, $< .05$.

Additional research is needed in which comparisons are made of member perceptions of group leaders from session to session. Comparison data should add to an understanding of the processes and movement of groups, since changes in member perceptions of group leaders could be related to patterns of leadership, as well as to phases through which groups move.

3. *The 44 laboratory members did not become more alike in the ways they saw the leaders as the laboratory continued: i.e., they did not vary less in their ratings of T-Group leaders in late compared to early sessions.* Results of t comparisons of variances in members' ratings of the group leaders in the six sessions revealed that the 44 members did in fact vary more—i.e., were less alike—in their ratings of the T-Group leaders for the fifth and sixth T-Group sessions than for either the first or the third session.[4]

The most obvious explanation of the increased dissimilarity in ratings of T-Group leaders by the 44 laboratory members might be the fact that four different leaders were being rated. If the laboratory experience did achieve the goal of increasing perceptual accuracy of participants, the increased variability could represent increased accuracy among members of each T Group, which would naturally result in increased difference between members of the four different groups, since they were rating four different leaders. Thus it is conceivable that while the 44 laboratory members did not become more alike in ratings of the four leaders, members within each separate T Group could nonetheless have become more alike in the ways they saw their group leader.

Such a trend would be in accordance with the Burke and Bennis (2) concept of perception congruity. According to this concept, group members are expected to reach greater agreement in their perceptions during the course of a group experience. Perception congruity may be determined by two operational measures: (*a*) congruence between a member's perceptions of himself and other members' perceptions of him, and (*b*) the degree of consensus among group members' perceptions of a particular member. The present question involves the second measure: i.e., the amount of consensus in members' perceptions of the group leader.

To determine then whether members within each of the four T Groups became more alike in their perceptions of the group leader, variances in members' ratings of the leader in the first and in the third T-Group session were compared with variances in the fifth and in the sixth session in each of the four T Groups. These were the same comparisons on which the 44 laboratory members as a whole were found to be less alike in their perceptions of T-Group leaders in the later session. According to results obtained in the comparisons within the separate groups, members in none of the four T Groups became more alike in the ways they saw the group leader. Moreover, members in one group became less alike in their perceptions of the group leader.[5]

---

[4] Data on the significant t comparisons are given below in the following order: sessions compared, respective SD, r between ratings for the two sessions compared, t, and p with $df = 42$.

1 and 5, 16.95 and 25.47, .10, 2.87, $< .01$.
1 and 6, 16.95 and 26.15, .35, 3.60, $< .001$.
3 and 5, 17.90 and 25.47, .48, 3.23, $< .01$.
3 and 6, 17.90 and 26.15, .42, 3.30, $< .01$.

Thus when members' early and late perceptions of the leaders were compared, results indicated that neither the 44 laboratory members nor the members within each of the four T Groups became more alike. No evidence was provided, therefore, to indicate that the laboratory experience led to greater consensus in members' perceptions of the group leaders.

4. *Compared to ratings for the first session, members of T Groups regarded the T-Group leaders more like the leaders regarded themselves in the last T-Group session.* Absolute differences between members' ratings of the leader and the leader's ratings of himself were computed, and an analysis of variance comparison revealed that the mean sum of member-leader differences on the 30 adjective scales for the last session was significantly smaller than the first session mean difference. (First and last session mean sums of member-leader differences were 44.07 and 37.66 respectively; $F = 9.57$; $df = 1,43$; $p < .01$.)

The finding that members came to see the leaders more as the leaders saw themselves is an instance of the first measure of perception congruity: i.e., congruence between a person's perception of himself and others' perceptions of him. Findings in this study support to some extent those of Burke and Bennis (2), even though different statistics were used in the two studies. While the present study used analysis of variance comparisons of mean absolute differences and t comparisons of correlated variances, Burke and Bennis used the statistically more sophisticated D scores and factor analytic procedures along with the Wilcoxon test. Their results indicated that (*a*) the way members regarded themselves and the way others regarded them became more similar, analogous to the present finding that the way leaders regarded themselves and members regarded them became more similar; but that (*b*) late in the laboratory, members varied as much as they had earlier in their perceptions of particular individuals on the Friendliness and Combination factors, analogous to the present finding that there was no difference between variances in members' perceptions of the group leaders early and late in the laboratory.[6]

---

[5] Data on the significant t comparisons in the one group are given below as in Footnote 4, with p given for $df = 9$.

1 and 6, 15.07 and 31.93, .66, 4.24, $< .01$.
3 and 6, 15.46 and 31.93, .75, 4.70, $< .01$.
3 and 5, 15.46 and 26.29, .75, 3.34, $< .01$.

It should be pointed out that although mean member ratings of the leaders were higher or more positive at the end compared to the beginning of the laboratory, this mean shift was not accompanied by a shift in the dispersion of individual members' ratings about the mean. Similarly, the mean shift toward smaller member-leader differences in perceptions of the leader was not accompanied by a shift toward smaller dispersion of members' ratings about their mean rating of the group leaders.

5. *Members changed their self-ratings in a positive direction from the beginning to the end of the human relations laboratory experience.* The difference between the mean total scores for initial and final self-ratings—means = 151.91 and 160.25 respectively—when tested for significance, yielded an F ratio of 9.76 which was significant beyond the .01 level ($df = 1,43$). This finding, along with the last two reported, should be regarded with caution in the absence of similar measures on a control group. Moreover, a possible halo affect, developing out of the increased involvement in the laboratory experience could influence responses in a positive direction. A replication, with the use of a control group, is needed to confirm this finding.

As a group, the 44 participants in this laboratory appeared to be homogeneous on several variables—age (most were 21), classification (mostly sophomores and juniors), grades (average of B or better), involvement in campus leadership activities, and a preference for the collegiate subculture. In a review of research on T Groups, Stock (7) reported that homogeneous National Training Laboratory groups at Bethel were seen as having less incentive to change, since member behavior was reinforced by a relatively stable culture. With regard to the present study, therefore, one question worth raising might be: Would the changes that appeared to be associated with the laboratory experience have been greater if the participants had been more heterogeneous on these demographic variables? The members could have been quite different on other variables, such as emotional and personality patterns, which are relevant to change. In fact, the magnitude of the standard deviation of their initial self-ratings, 17.4, suggests that they were rather heterogeneous with respect to self-perceptions and possibly their self-concepts at the outset of the laboratory. The influences of homogeneity *versus* heterogeneity on group development, on incentives for behavior change, and on resistance to change, though not investigated in this study, represent an important area for T-Group research. Harrison (4) predicts greater learning from the conflicts which grow out of polarized subgroups within a heterogeneous group, in contrast with a more homogeneous group in which conflict is less likely to occur.

---

[6] In contrast, however, Burke and Bennis (2) also found that members did come to agree more with regard to extent of participation of particular individuals.

6. *T-Group members changed their self-ratings on 11 of the 30 adjective scales.* In their final, compared to initial, self-ratings, they rated themselves as significantly more warm, trusting, sensitive, supporting, deep, caring, authentic, relaxed, easy, natural, and dependent. (F ratios ranged from 4.76 to 12.77; $df = 1{,}43$; $p < .01$ for six comparisons and $p < .05$ for five comparisons.)

On the first seven of these 11 scales, the mean was toward the positive adjective initially and became more positive in final self-ratings. In contrast, on the next three scales—tense-relaxed, anxious-easy, and self-conscious-natural—the mean initial self-rating by T-Group members was toward the negative adjective (listed first) and shifted across the mid-point (4.0) toward the positive adjective in final ratings. On one adjective scale—independent-dependent—T-Group members saw themselves as changing in the "negative" (dependent) direction, although the mean final self-rating was still toward the positive adjective. This shift could represent a positive movement toward the development of interdependent attitudes and behaviors.

These positive adjectives, though limited by the design of the research instrument, describe to some extent the kinds of member behavior regarded as desirable outcomes of the group experience. At the same time the emphasis was not simply on achieving such specific behavior outcomes. Rather, the main focus was on the process by which members and leader experienced and learned from their encounters with each other. These adjectives also represent partial descriptions of the developing group climate or culture which is in process and never completely achieved.

7. *The 44 laboratory members were not more alike in the ways they saw themselves at the end compared to the beginning of the group experience,* according to the t test of the difference between correlated variances. (*SD*s for initial and final total self-rating scores were 17.24 and 17.95 respectively, $r_{1.2} = .49$, $t < 1.0$.) Thus the mean positive change in members' self-ratings was not accompanied by a change in the dispersion of individual members' ratings about the mean.

Although the t test revealed that members in the total sample did not vary less and, hence, did not become more alike in their overall self-ratings, the question was raised as to whether members of the same T Group became more like one another during the course of the group experience. It is conceivable that pressures toward conformity, if they do emerge during the laboratory experience, might be observed empirically in the separate T Groups, but not in the laboratory group overall. Consequently, the variances in members' initial and final self-ratings were compared separately for each of the four T Groups. None of the four *t* values approached statistical significance. Thus, as was true for the total sample of laboratory members, members of the same T Group were not more alike in the ways they saw themselves at the end compared to the beginning of the group experience.

The element of heterogeneity suggested by this finding, though not antithetical to group cohesiveness, appears somewhat at odds with the theoretical notion that group members tend to become more conforming. Changes in group variability might be unexpected since, as pointed out earlier, the members appeared to be rather homogeneous demographically in several ways. Still, homogeneity on demographic characteristics did not necessarily restrict heterogeneity with respect to emotionality and personality patterns. The magnitude of the two standard deviations above indicates that there was in fact considerable variability in the self-images of members at the outset and that it was possible for them to become more similar, as would be shown if there were less variability at the end of the laboratory.

D. SUMMARY

Bipolar adjective ratings by 44 university students and by their leaders in four T Groups were analyzed for information on how members of a human relations laboratory change in their perceptions of their group leader and of themselves. To measure change in members' perceptions of leader and self, data obtained early and late in the laboratory were compared. The statistical data included (*a*) mean number ratings of leader and of self, (*b*) variances in members' ratings of leader and of self, and (*c*) mean absolute differences between each member and his leader in ratings of self and of leader. The statistical methods included the analysis of variance with single classification and repeated measurements, and the *t* test of the difference between correlated variances.

Contrary to expectations, T-Group members' self-ratings did not become more like self-ratings of their leaders at the end of the group experience. Analysis of members' ratings of leaders, however, indicated (*a*) that members saw leaders more positively in the last T-Group session than in the first session, but (*b*) that the shifts in members' perceptions of the leaders from session to session were not continuously positive.

The 44 members as a whole and the members in the separate T Groups did not become more alike in their perceptions of the group leaders at the end, compared to the beginning, of the group experience. It was found, however, that members came to perceive the leaders more like the leaders perceived themselves.

T-Group members saw themselves more positively at the end than at the beginning of the laboratory experience. In essence, they regarded themselves as being more authentic and natural; more sensitive and supporting; and more able to trust and to care for others and themselves. The significant positive change in the mean of members' self-ratings, however, was not accompanied by a tendency for group members to become more similar in their self-ratings. There was considerable variability in the self-perceptions of members at the outset of the laboratory experience. Even though it was possible, then, for members to become more similar during the experience, comparison of variances in their initial and final self-ratings—both for the 44 laboratory members as a whole and for members of each of the four T Groups—yielded no significant difference. Thus, parallel to the finding that the 44 members as a whole and the members of separate T Groups did not become more alike in their perceptions of the group leader was the finding that these members did not become more alike in their ratings of self.

REFERENCES

1. BRYANT, R. J. Changes in perceptions of self and leader in small groups. Unpublished Master's thesis, The University of Texas at Austin, 1967.
2. BURKE, R. L., & BENNIS, W. G. Changes in perception of self and others during human relations training. *Hum. Relat.*, 1961, **14**, 165-182.
3. FLETCHER, J. R. An analysis of developmental groups in a human relations laboratory. Unpublished Master's thesis, The University of Texas at Austin, 1967.
4. HARRISON, R. Problems in the design and interpretation of research on human relations training. In *Explorations: Human Relations Training and Research, No. 1.* NTL Institute for Applied Behavioral Science, Associated with the National Education Association, 1201 Sixteenth Street, N.W., Washington, D.C., 1967.
5. OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. The Measurement of Meaning. Urbana, Ill.: Univ. Illinois Press, 1957.
6. PETERS, D. R. Identification and personal change in laboratory training groups. Doctoral dissertation, Alfred P. Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Mass.: Microreproduction Laboratory, M.I.T. Libraries, 1966.
7. STOCK, D. A survey of research on T Groups. In L. P. Bradford, J. R. Gibb, & K. D. Beene (Eds.), *T-Group Theory and Laboratory Method.* New York: Wiley, 1964. Pp. 395-441.

Department of Educational Psychology
The University of Texas at Austin
Austin, Texas 78712

Department of Counseling
The University of Maryland
College Park, Maryland 20742

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 21-31.

THE RELATIONSHIP OF ILLNESS AND HEALTH BEHAVIOR TO PAPER-AND-PENCIL TESTS*

Bolt Beranek and Newman, Van Nuys, California

JEAN S. KERRICK AND VIRGINIA A. CLARK[1]

A. INTRODUCTION

The University of California at Los Angeles (UCLA) Peace Corps study began as a study of the effectiveness of health training. The usual paper-and-pencil tests were taken before and after health training; and changes in knowledge and attitudes toward illnesses were assessed (6).

The investigators had anticipated before-after changes in the paper-and-pencil tests. They believed, however, that a problem of greater significance was the problem of actual behavior change which would reduce morbidity. Peace Corps volunteers moving to countries with far different health problems must adopt radical new health practices if they are to avoid unfamiliar and unpleasant illnesses.

The problems involved in exploring behavior changes and morbidity reduction in this study were numerous. The major ones follow:

(*a*) In the few studies where attitudes and behavior have been jointly studied, no very strong relationships have been found (3). Perhaps the lack of significant associations between attitudes and behavior arises from our inability to measure the motivating attitudes. (The student dislikes examinations but he takes them because he has a favorable attitude toward staying in school.)

(*b*) If it is assumed that the relevant attitudes can be measured, the appropriate behaviors cannot be. In controlled studies, limited single-event behaviors may be measured, but complex ones cannot. Even the simple behavior must be narrowly defined before it can be measured in any practical sense. A visit to a college health clinic can be verified, but a visit to a private physician (during

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 31, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This study was supported by Peace Corps Contract PC-(W)-381. Computing assistance was obtained from the Health Sciences Computing Facility, University of California at Los Angeles, sponsored by NIH Grant FR-3. The authors would like to thank Dean L. S. Goerke and Dr. Donald T. Rice for their assistance in early phases of this study. Dr. Clark is Assistant Professor of Biostatistics at the University of California at Los Angeles. Dr. Kerrick is now at Bolt Beranek and Newman Inc., Los Angeles.

the summer, for example, when large numbers of students move to large numbers of residences) cannot be verified. Trained observers cannot agree upon the beginning and ending of a complex behavioral interaction. Even if behavioral measurement were possible, it would not be practical to attempt measurement of behaviors overseas for a two-year span for several hundred Peace Corps volunteers.

(*c*) The measurement of morbidity is often imprecise.[2] Self-report has consistently produced underestimates of at least some kinds of illness (1, 7). Physician visits, even where they can be determined, reflect only those illnesses for which the individual believes he needs or wants treatment. Data from the present study demonstrate individual differences in the extent to which symptoms are judged to warrant a physician visit.

(*d*) Health behavior and morbidity measures may be the same or they may not. A physician visit may be an index of illness or morbidity; but the conscientious Peace Corps volunteer may make physician visits to preserve good health or to insure early diagnosis and treatment, so that physician visits become a measure of good health behavior.

With full awareness of these difficulties, the investigators proceeded with the assumption that useful data could be collected to begin the task of relating illness and health behavior to paper-and-pencil tests of attitudes and information. The measure of illness and health behavior used was physician reports, which are filled out when a volunteer visits him. It was assumed and later verified that many illnesses among the volunteers were closely related to health behavior. That relationship suggested morbidity as a negative measure of behavior. In the morbidity reports prepared by Peace Corps physicians, the incidence of amebiasis, dysentery, and other gastro-enteric diseases was sizable and was exceeded only by colds or flu. This high incidence is indication of a violation of recommended practices in preparation and ingestion of food and drink. As an extreme but clear example arguing for the use of morbidity measures as indices of health behavior, consider the incidence of venereal disease: the absence of illness does not tell us that certain behaviors did *not* take place, but the presence of the illness tells us that it *did*.

Given crude morbidity measures and a battery of psychological tests, the investigators offer some hypotheses relating health behavior and illnesses to attitudes, and suggest more or less promising avenues for further investigations.

---

[2] Initially it was felt that valid and reliable morbidity indices would be obtained by examination of stool specimens and of sera. Unfortunately, analysis of the first set of stool specimens produced such a low frequency of positive laboratory results that these data were not analyzed. Similarly, of the first 63 serum samples analyzed by the Communicable Disease Center (CDC), Atlanta, Georgia, only one showed a disease which might have been contracted in Africa; so that CDC does not plan to analyze the remaining specimens.

B. METHOD AND PROCEDURE

Peace Corps volunteers trained at a single university and headed for a single African country were given a battery of tests at the beginning of training (Exam I), at the end of training (Exam II), after six to eight months in the field (Exam III), and at the end of two-year service (Exam IV). Peace Corps physicians' Monthly Morbidity Reports were examined for all volunteers. Step-wise discriminant functions were used to relate the paper-and-pencil tests to the crude morbidity measures.

Before and after training, a health information test of 24 multiple-choice items and 10 matching questions was given. Most of the items had been previously used in Peace Corps health training courses. At the end of the two-year service period 10 relatively common host country illnesses were listed and the volunteer was asked the signs and symptoms of each illness.

A previously constructed form of the semantic differential (5) was used before and after training and in the field, to assess two major dimensions of judgment toward illnesses. The first dimension, "seriousness or severity," is attitudinal in nature, and the second major dimension measures belief about the "avoidability" of illnesses. The initial tests (before and after training) measured these two dimensions for 25 selected illnesses; the field test was shortened to 11 representative illnesses.

A special form of the semantic differential was constructed and pretested for use in the present study to indicate the likelihood with which volunteers expect to follow major recommended health related behaviors. The scale was not developed as a general measure of attitude toward behavior, nor as a general description of the means by which behaviors are differentiated. It was constructed solely to determine the expressed likelihood[3] of following certain practices. Twenty-seven behaviors were included in the tests given during the training period; 12 of the 27 were included in the field tests.

At the end of training, a course reaction sheet of the traditional closed-end type was administered; for example, students were asked if they found the course informative, clear, interesting, and so forth.

Physicians' Monthly Morbidity Reports from the field were examined for number of physician visits, number of hospitalizations, number of illnesses with a psychological component, and number of days lost from work. While it is possible that some volunteers received treatment from non-Peace Corps physicians, those visits must be reported to the Peace Corps before bills would be paid. Thus, it is likely that the records include the majority of physician visits during the service period.

---

[3] Computation of scale scores is explained by Clark and Kerrick (2).

C. Results

For (*a*) the total number of physician visits, (*b*) the days which the volunteer missed work because of illness, (*c*) the days the volunteer was hospitalized, and (*d*) the presence or absence of an illness with any psychological component, two sets of step-wise discriminate functions were performed for each index. Essentially, the questions asked were these: first, can the number of physician visits, days missed work, days hospitalized, or psychological illness be predicted from the paper-and-pencil tests given the volunteers before and after training; and second, in view of the changes in attitudes and expected behaviors over field-service period, can better prediction of these morbidity variables be achieved by using all paper-and-pencil tests including those taken overseas?

1. *Prediction of the Number of Physician Visits*

Table 1 gives the best prediction achieved for extreme groups of volunteers in regard to physician visits, the variables for prediction, and the *F* level or significance of that prediction. With only tests administered in this country before the volunteer left for overseas service, three tests give us the best prediction of whether or not a volunteer will have a large number of physician visits. Those with the most visits showed a higher personal health likelihood score on Exam I (pretraining), they found illnesses more avoidable after training, and they also showed a higher personal health likelihood score on Exam II (post training) than did those with fewer visits.

The proportion of correct placement into one of the two groups was not

TABLE 1
Physician Visits Predicted by Three Tests and Physician Visits Reported
(Exams I and II only)

Number visits reported	Number visits predicted[a]		
	0-3	4+	Total
0-3	70	46	116
4+	16	29	45
Total	86	75	161

Note: Percent correctly predicted: 61.5; $F = 2.10$; $df = 3,157$; $p < .10$.

[a] Predictive tests: (*a*) Likelihood, Personal Health, Pretraining; (*b*) Avoidability of Illness, Posttraining; (*c*) Likelihood, Personal Health, Posttraining.

high. Roughly 61 percent of the volunteers were correctly placed in one of the two groups, with about 30 percent false negatives, and about 40 percent false positives. Thus, close to 40 percent of those whom the authors felt would have any physician visits, actually did not, and about 40 percent of those whom the authors felt would have few physician visits, actually had many; yet the F ratio approaches significance.

When all tests are considered, no improved prediction is achieved. When the grossness of the measures and the time period are considered, it is surprising that any relationship at all is found between tests given in this country, and physician visits during the two-year service visits overseas. It is interesting to note, however, that speculations made from summer observers' reports appear justified. Physician visits should not be considered as a measure of morbidity alone, but should be considered partially as a measure of good health behavior. The volunteer, who, in this country, says that he is most likely to follow personal health behaviors also appears most likely to visit a physician when he is ill.

Further evidence of this confounding of morbidity with good health behavior is obtained by a closer examination of the scores on likelihood and health behavior. This scale tends to be related in curvilinear, rather than linear fashion to each of the morbidity measures. Thus, if volunteers are separated into three groups rather than two (those with no physician visits, an average number of physician visits, and a great many physician visits), volunteers are found who express a low likelihood of following recommended behaviors in both extremes. The volunteer unlikely to follow recommended personal behaviors tends either not to visit a physician or has so many illnesses that he shows many physician visits. (It should be noted that the discriminant function is a linear function. The pronounced curvilinearity found especially in the likelihood scale decreases the ability of the investigators to predict adequately the number of physician visits. It is possible that, if number and kind of physician visits were recorded and analyzed separately according to the presumed behavioral cause, better prediction would result.)

2. *The Prediction of the Number of Workdays Missed*

By and large, when volunteers reported missing work because of illness, they were hospitalized. The present section deals with the small group of volunteers who reported missing work because of illness when they were not hospitalized.

Four variables, extracted by step-wise discriminant function analysis, enable us to predict correctly 67 percent, with almost 30 percent false negatives and a little over 30 percent false positives (Table 2). Again, this prediction is based initially on tests given at the training sites. Those variables entering into the prediction were posttraining avoidability scores (those who missed work felt illnesses were much more avoidable) the second and first knowledge tests (where volunteers who reported missing work initially scored lower on the knowledge test, but after training scored higher on the knowledge test) and finally, Guttman scores derived from judgments of the personal health section of the health training (those missing work expressed greater dislike for the health training section). Again, the *F* ratio approaches significance, falling beyond the .10 level.

TABLE 2
DAYS MISSED WORK PREDICTED BY FOUR TESTS AND DAYS REPORTED
(Exams I and II only)

Reported days missed	Predicted days missed[a] None	Some	Total
None	98	49	147
Some	4	10	14
Total	102	59	161

Note: Percent correctly predicted: 67.1; $F = 2.25$; $df = 4{,}156$; $p < .10$.
[a] Predictive tests: (*a*) Avoidability of Illness, Posttraining; (*b*) Knowledge, Posttraining; (*c*) Knowledge, Pretraining; (*d*) Course Reaction.

With the use of all paper tests, prediction of those who will report missing work because of illness is not greatly increased, but the same level of prediction can be reached by using only three variables. The first variable in prediction remains the judgment of the avoidability of illness on Exam II (posttraining); the second variable indicates that at the end of training those who reported missing work also reported a much higher likelihood of following recommended personal health behaviors, and a somewhat lower knowledge of frequent symptoms of illnesses at the end of service knowledge test. It would appear that responses to the end of service questions about the likelihood of following personal health may well be the results of experience with illness, rather than a causative factor in determining who will be ill enough to miss work. Those volunteers who, during their overseas service, have been forced to miss work because of illness, even though that illness was not severe enough to cause hospitalization, are perhaps better able to appreciate the necessity for following recommended health-related behaviors. On the other hand, these volunteers know fewer of the symptoms which accompany the illnesses in the host country (4).

After a review is made of the training test scores which predict the days missed work in relationship to those scores which predict physician visits, it is interesting to note that the single scale in common with the two is the avoidability score on Test II (posttraining), where both those with a higher number of physician visits and those who miss work see illness as more avoidable than a group with an average number of physician visits and the group who failed to report days missing work because of illness. It may be argued, of course, that many more volunteers missed work because of illness than actually reported it; and it would seem likely that some volunteers did miss some work and fail to report it. It is impossible from our data, however, to assess the degree of underreporting in this area. It seems safe to say, however, that those who reported missing work because of illness actually did miss a full day of work on each occasion that they reported it. In this respect again, ability to predict may be limited by the accuracy of reporting, and this perhaps reflected in the greater number of false positives than in the number of false negatives. It is quite possible that some of those whom the authors predicted would miss work actually did miss work because of illness but simply failed to report it.

3. *The Prediction of Days Hospitalized*

The *F* ratio failed to reach significance at the .10 level (Table 3) for the prediction of days hospitalized. In the step-wise analysis, adding any variable past the first simply increases the false negatives. Although this analysis did not produce an *F* level even approaching significance, it is instructive to look at the variables extracted.

Of those tests taken before the volunteers left this country, the first three variables extracted by analysis are Exam I, likelihood of personal health behavior; Exam II, avoidability; and Exam II, likelihood of personal behavior. Those who had extensive hospitalization showed a much lower likelihood of following recommended personal health behaviors than did those without

TABLE 3

DAYS HOSPITALIZED PREDICTED BY ONE VARIABLE AND THOSE REPORTED
(Exams I and II only)

Hospitalization reported	Hospitalization predicted[a] None	Some	Total
None	76	57	133
Some	11	16	27
Total	87	73	150

Note: Percent correctly predicted: 61.3; $F = 1.80$; $df = 1{,}148$; $p > .10$.

[a] Predictive tests: (*a*) Likelihood of Intended Health Behavior, Pretraining; (*b*) Avoidability of Illness, Posttraining; (*c*) Likelihood of Intended Health Behavior, Posttraining.

extensive hospitalization for both Exam I and Exam II. Those with extensive hospitalizations did report increased likelihood of following recommended personal health behaviors after training than before, but at both pre- and post-training tests they were well below the average likelihood for the group which did not have extensive hospitalization. Also, although the difference was small, those with hospitalization had a slight tendency to find illnesses more avoidable.

Adding all tests, including those taken in the host country, fails to provide any improved prediction of those who will have extensive hospitalization. The same three tests—Exam I, personal health; Exam II, avoidability; and Exam II, personal health—are extracted by analysis. Examination of results of the full set of tests, however, suggests that extensive hospitalization does produce change itself. Thus, the group with extensive hospitalization reports a slightly higher likelihood of following recommended personal health behaviors at Exam period IV and at the end of training.

It is interesting to see that the number of physician visits and the days hospitalized are differentiated by the same tests. For both morbidity measures, the first tests found important were extracted in the same order. In general then, it might be said that those volunteers who express a high likelihood of following recommended personal health behaviors on both Exams I and II will show a higher number of physician visits but a lower number hospitalizations. (Again, the same curvilinearity of the personal health behavior scale probably decreases ability to predict correctly the days hospitalized. However, the prediction rate remains at about 60 percent correctly predicted.)

4. *Prediction of Volunteers with Any Illness Having a Psychological Component*

It must be admitted that this scoring category is probably the least reliable. The rating of the psychological component of illness is dependent on physician judgment; it is, therefore, susceptible to interphysician variability. Nonetheless, with the realization that the work was conducted with very gross measures of morbidity, tests of those 57 volunteers who showed any illness which was rated as having a psychological component on the physicians' Monthly Morbidity Report were examined statistically. With reference to only those tests administered in this country, three variables predict with almost 60 percent accuracy, the F ratio falling just short of .10 (Table 4). Adding all exams fails to increase predictability appreciably, although it does decrease the number of false positives: that is, those who are predicted to have illnesses with psychological components, but who indeed do not. When analysis is

TABLE 4
NUMBER OF VOLUNTEERS HAVING AN ILLNESS WITH ANY PSYCHOLOGICAL COMPONENT, PREDICTED BY THREE VARIABLES
(Exams I and II only)

Reported illnesses with psychological component	Predicted illnesses with psychological component[a]		
	None	Some	Total
None	77	61	138
Some	28	44	72
Total	105	105	210[b]

Note: Percent correctly predicted: 57.6; $F = 2.03$; $df = 3{,}206$; $p > .10$.

[a] Predictive tests: (*a*) Severity of Illness, Pretraining; (*b*) [Knowledge, Posttraining (not used in prediction)]; (*c*) [Knowledge, Pretraining (not used in prediction)].

[b] Number includes volunteers who left before the end of service.

limited to tests given in this country or when analysis is extended to all tests, it is found that prediction of psychologically related illnesses depends primarily upon attitudes toward illness and upon knowledge.

With reference to tests given in this country, those who eventually have illnesses with psychological components find illnesses more severe than do subjects not having illnesses with psychological components. They know slightly more about illness at the beginning of training and somewhat less at the end, as measured by knowledge tests I and II. Additionally, they find illnesses more avoidable at the beginning of training. When all tests are considered, the most important difference between the two groups is in the knowledge test given at the end of training; the group *with* any psychological component to their illnesses knows more illness symptoms than does the group with no illnesses having psychological components. While the group having illness with a psychological component originally found illness more avoidable, at the end of service examination they found it less avoidable than did those without similar psychological correlatives to illness.

Of the four morbidity indices used, two of them—the days of work missed due to illness and the presence of an illness with a psychological component—appear to be related to judgments of illness and to knowledge. The other two—the number of physician visits and extensive hospitalization—appear more strongly related to the expectation of following recommended personal health behaviors.

D. SUMMARY AND DISCUSSION

Paper-and-pencil tests taken in this country before Peace Corps volunteers left for a two-year service period in Africa were related to four relatively crude measures of morbidity: (*a*) number of physician visits, (*b*) days missed work because of illness (without hospitalization), (*c*) hospitalization, and (*d*) illness rated by the Peace Corps physician as having a psychological component.

Attitude toward illness (or judgment of the severity of the illnesses common in the host country) was related only to the opinion of the physician that the volunteer had some illness with a psychological component. A very tentative suggestion may be made that attitude toward illness may be predictive of psychosomatic illness or to somatizing of work or living pressures rather than to more definite physical illness.

Belief in the avoidability of illness, however, is related to all of the other three measures of morbidity. The scale of likelihood of following recommended personal health behaviors is related both to the number of physician visits and the hospitalization for illness, in conjunction with judgments of avoidability. The relationships are not simple ones, and the present data indicate a need for greater refinement and precision in the measurement of morbidity and in the paper-and-pencil tests used.

Because so many variables other than attitudes, information, and intended health behavior may be important; because of the gross categorization of illness and morbidity; and because of the two-year time span, it is surprising indeed to find that the tests related to the authors' morbidity measures and gave some indications of the volunteer's health-related behavior in the field.

Health behavior, itself, is an intervening variable in only some instances, and illness will result only occasionally when poor health practices are followed. Nonetheless, the number of physician visits (very few *vs.* many) and the days hospitalized because of illness (none or many) are highly related to scores of the personal health behavior tests, especially the tests given prior to service abroad. These same two gross morbidity measures are related to judgments of the avoidability of illness which the volunteer makes immediately following training.

Future investigations might well focus upon more refined measurements, and upon the complex interaction between health behavior, illness experience, and subsequent health behavior.

Implicit in the present data is the proposition that some people learn from training or course work, resolve to follow recommended practice, and, in general, remain healthy. Others, conversely, who felt themselves ill enough to miss work (although not so ill that they required hospitalization) learned little of the signs and symptoms of illness, but ended service with greater verbal resolution to follow recommended practices. For this latter group, we might well ask whether there is some way to teach them to recognize real illness, and some relatively easy way to convince them to take reasonable precautions to preserve their health.

REFERENCES

1. BELLOC, N. B. Validation of morbidity survey data by comparison with hospital records. *J. Amer. Stat. Assoc.,* 1954, **49**, 832-846.
2. CLARK, V. A., & KERRICK, J. S. A method of obtaining summary scores from semantic differential data. *J. of Psychol.,* 1967, **66**, 77-85.
3. FESTINGER, L. Behavioral support for opinion change. *Public Opin. Quart.,* 1964, **28**, 404-417.
4. GOERKE, L. S., RICE, D. T., KERRICK, J. S., & CLARK, V. A. Research in the effectiveness of methods of education in training for health among Peace Corps volunteers. Final Report, Peace Corps Contract PC-(w)-381, Los Angeles, California, 1967.
5. KERRICK, J. S. The use of the semantic differential and its special application to illness. *Health Educ. Monog.,* in press.
6. KERRICK, J. S., CLARK, V. A., & RICE, D. T. Lecture versus participation in the health training of Peace Corps volunteers. *J. Educ. Psychol.,* 1967, **58**, 259-265.
7. SAGEN, O. K., DUNHAM, R. E., & SIMMONS, O. Health statistics from record sources and household interviews compared. Social Statistics Section, American Statistical Association, Washington, D.C., December, 1959.

Bolt Beranek and Newman Inc.
Consulting Development Research
15808 Wyandotte Street
Van Nuys, California 91406

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 33-34.

A NOTE ON THE USE OF DISCREPANCY SCORES IN A SELF-CONCEPT INVENTORY*

Department of Psychology, University of Manitoba

SHIRIN SCHLUDERMANN AND EDUARD SCHLUDERMANN

Factor analyses of self- and ideal-ratings on Worchel's self-concept inventory (2) failed to yield meaningful unidimensions for the replicated sample of college boys and girls (1). It was considered necessary, therefore, as the next logical step, to examine carefully the discrepancy scores between the self- and ideal-ratings with a view to obtain basic dimensions which would account for maximum variance in discrepancy scores. Since the discrepancy scores are taken as indices of adjustment or maladjustment, it is all the more important to clarify what is involved in the discrepancy measures. Discrepancy scores were obtained on each item between the self- and ideal-scores for each *S*. Intercorrelations among discrepancy scores on 54 × 54 items were obtained for boys and girls separately. Two sets of factor matrices were obtained from two sets of correlation matrices. Factoring was done by the principal axes method and rotated orthogonally by the varimax solution. The first factor accounted for 16 and 12 percent of variance respectively, in boys' and girls' factor matrices. A total of seven factors in the boys' case and eight factors in the girls' case accounted for approximately 40 percent of the total variance. As is clear, there are no common variances; hence the discrepancy scores consist of a large number of specific factors.

To obtain the relative contributions of the ideal-ratings and self-ratings to the discrepancy scores, product-moment-correlation coefficients were computed for boys and girls separately (see Table 1).

The results agree with Wylie's statements (3, p. 105) that self-scores contribute most of the variances to the discrepancy scores. However, there are interesting differences between the sex groups; the correlation between self- and ideal-scores is near zero for girls, whereas the correlation for boys is .33. In the girls' case, the correlation between self *vs.* discrepancy scores was found to be significantly higher than the correlation between ideal *vs.* discrepancy scores. In the boys' case, the correlation of .62 was not found to be significantly different from a correlation of .46. The analyses reported here indicate that

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 6, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 33-34.

A NOTE ON THE USE OF DISCREPANCY SCORES IN A SELF-CONCEPT INVENTORY*

Department of Psychology, University of Manitoba

SHIRIN SCHLUDERMANN AND EDUARD SCHLUDERMANN

Factor analyses of self- and ideal-ratings on Worchel's self-concept inventory (2) failed to yield meaningful unidimensions for the replicated sample of college boys and girls (1). It was considered necessary, therefore, as the next logical step, to examine carefully the discrepancy scores between the self- and ideal-ratings with a view to obtain basic dimensions which would account for maximum variance in discrepancy scores. Since the discrepancy scores are taken as indices of adjustment or maladjustment, it is all the more important to clarify what is involved in the discrepancy measures. Discrepancy scores were obtained on each item between the self- and ideal-scores for each *S*. Intercorrelations among discrepancy scores on 54 × 54 items were obtained for boys and girls separately. Two sets of factor matrices were obtained from two sets of correlation matrices. Factoring was done by the principal axes method and rotated orthogonally by the varimax solution. The first factor accounted for 16 and 12 percent of variance respectively, in boys' and girls' factor matrices. A total of seven factors in the boys' case and eight factors in the girls' case accounted for approximately 40 percent of the total variance. As is clear, there are no common variances; hence the discrepancy scores consist of a large number of specific factors.

To obtain the relative contributions of the ideal-ratings and self-ratings to the discrepancy scores, product-moment-correlation coefficients were computed for boys and girls separately (see Table 1).

The results agree with Wylie's statements (3, p. 105) that self-scores contribute most of the variances to the discrepancy scores. However, there are interesting differences between the sex groups; the correlation between self- and ideal-scores is near zero for girls, whereas the correlation for boys is .33. In the girls' case, the correlation between self *vs.* discrepancy scores was found to be significantly higher than the correlation between ideal *vs.* discrepancy scores. In the boys' case, the correlation of .62 was not found to be significantly different from a correlation of .46. The analyses reported here indicate that

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 6, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

TABLE 1
PRODUCT-MOMENT-CORRELATION COEFFICIENTS BETWEEN IDEAL SCORES, SELF-SCORES, AND DISCREPANCY BETWEEN IDEAL- AND SELF-SCORES

	Replicated sample		Normative sample
Scores	Boys ($N = 177$)	Girls ($N = 150$)	Boys ($N = 153$)
Ideal *vs.* Self	.33	.05	.49
Ideal *vs.* Discrepancy	.46*	.30*	
Self *vs.* Discrepancy	.62*	.60*	

* Significant at .01 level.

Worchel's self-concept inventory is neither a unidimensional nor a multidimensional scale. However, the overall composite scores on self and ideal and the discrepancy scores may be positively correlated to other tests.

REFERENCES

1. SCHLUDERMANN, S., & SCHLUDERMANN, E. A methodological note on "A Self-Concept Inventory." *J. of Psychol.*, 1969, **71**, 259-260.
2. WORCHEL, P. Adaptability-screening of flying personnel: Development of a self-concept inventory for predicting maladjustment. School of Aviation Medicine, Report No. 56-62, U.S.A.F., Randolph Air Force Base, Texas, 1957.
3. WYLIE, R. C. The Self-Concept: A Critical Survey of Pertinent Research Literature. Lincoln: Univ. Nebraska Press, 1961.

Department of Psychology
University of Manitoba
Winnipeg, Canada

Published as a separate and in *The Journal of Psychology,* 1969, **72**, 35-40.

PREDICTION OF SUPERVISORY SUCCESS BY USE OF STANDARD PSYCHOLOGICAL TESTS*

Department of Psychology, Muskingum College

H. EDWIN TITUS[1]

A. INTRODUCTION

Widespread use of psychological tests for personnel appraisals in selection, placement, and promotion considerations is a fact of life in industry today. Barrett (1) has pointed out that appraisal of the man without specification of his work situation is a "patent medicine" approach and is held in low esteem "for the simple reason that it is unlikely to work" (1, p. 141). Test-based appraisals are used in a variety of ways some of which are legitimate and sophisticated. Sadly, however, many applications are highly questionable. Within the limitation of a clear definition of a specific situation, both the statistical and clinical interpretation of test data can produce some reasonable forecasts of adjustment, stability, adaptability, and aptitude. We see increasingly often, however, forecasts being made well beyond what can be justified by the available data concerning the predictive validity of the devices that are employed. It is a serious problem when "logical" or *a priori* intervening connections between predictors and criteria are simply assumed on the basis of old simplistic models, despite the evidence showing that such an approach is inadequate (cf., 2, 3).

It may seem reasonable, for example, that a supervisor who tests higher on "ascendance" will be more successful than one who tests lower. As an hypothesis to be tested under particular conditions in a carefully defined setting, such a proposition is legitimate and possibly useful. However, to accept and act upon such a proposition without validation specific to the situation is likely to lead to grossly erroneous results.

The present report provides data on the relationship between certain psychological tests that have often been involved in industrial appraisals and a diversified set of criteria for supervisory success. The purpose here is not to say that some tests are useful and others are not, but rather to provide data

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 6, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The author wishes to thank Dr. Robert Krug for his valuable advice and suggestions.

showing the patterns of relationship of a substantial group of predictors against a diversified set of criteria in a common setting with all data based on a homogeneous sample of line supervisors. Such data can provide a starting point for examining measures that may be considered in specific local studies of personal factors in supervision.

B. METHOD

As a part of a larger investigation, nine tests yielding 38 scores were administered to 67 line supervisors at a large U.S. Government supply depot. Participation in the study was voluntary, but strongly encouraged by higher management. Those who participated were given time off from their regular duties and were tested in special testing sessions which took place during regular working hours. Participants were assured that the study was purely for research purposes and not a basis for administrative action. The vast majority of the middle-level supervisors participated; however, for various reasons it was not possible to obtain complete data on a few of those who were willing to participate. The 67 subjects mentioned above included only those for whom the data were complete in all categories.

The battery of standard psychological tests included The American Council on Education Psychological Examination (ACE), two variables; The California F Scale; The Empathy Test; The Hidden Figures Test; The Guilford-Zimmerman Temperament Survey (G-Z), 10 variables; How Supervise?; The Sixteen Personality Factor Questionnaire (16 PF), 16 variables; The George Washington University Social Intelligence Test (SIT), five variables; and The Watson-Glazer Critical Thinking Appraisal (W-G).

The criterion measures were chosen on the basis of the availability of information and the interests of higher management. Since any modern consideration of industrial validity of psychological tests must recognize that the day of seeking the "ultimate criterion" is long past (cf., 2), it was considered essential to have as broad a spectrum of criterion measures as possible. The criterion measures included four morale variables, six measures of the performance of each supervisor's work group,[2] and a proficiency rating for each supervisor provided by their superiors. The specific variables were as follows: morale concerning working conditions and environment (M-1), morale concerning financial reward (M-2), morale concerning supervision (M-3), morale concerning management and administration (M-4), average number of accidents (P-1), average number of suggestions submitted (P-2), average number of awards received for suggestions (P-3), average number of subordinates receiving awards for superior performance (P-4), average number of hours lost due to sick leave (P-5), average number of hours lost due to unauthorized absence (P-6), and proficiency rating of supervisors (R-1).

---

[2] In a service facility like the present research setting, production measures of performance in the conventional sense were, of course, not possible.

C. RESULTS

A list of the tests and the variables which they purport to measure is given in Table 1 along with a code designation, mean, and standard deviation for

TABLE 1
CODE DESIGNATIONS, MEANS, AND STANDARD DEVIATIONS OF THE TEST VARIABLES

Code	Test	Variable name	Mean	Standard deviation
T-1	ACE	Linguistic Ability	41.64	15.57
T-2	ACE	Quantitative Ability	20.14	11.08
T-3	F Scale	Authoritarianism	134.61	23.65
T-4	Empathy	Mass Empathy	72.30	19.14
T-5	HFT	Percep. Reorg.	5.18	2.83
T-6	G-Z	Activity	16.33	5.37
T-7	G-Z	Restraint	17.22	4.57
T-8	G-Z	Ascendance	15.12	4.72
T-9	G-Z	Social Interest	18.20	5.74
T-10	G-Z	Emotional Stability	18.93	4.76
T-11	G-Z	Objectivity	17.16	5.22
T-12	G-Z	Friendliness	14.79	5.85
T-13	G-Z	Reflectiveness	18.78	4.19
T-14	G-Z	Cooperativeness	17.54	5.08
T-15	G-Z	Masculinity	17.88	4.21
T-16	How Sup.?	Supervisory Knowl.	41.42	11.11
T-17	16 PF	Cyclothymia	8.36	2.72
T-18	16 PF	Intelligence	6.01	1.92
T-19	16 PF	Ego Strength	15.08	2.97
T-20	16 PF	Dominance	12.50	3.22
T-21	16 PF	Surgency	12.71	3.16
T-22	16 PF	Super-Ego Strength	13.08	2.62
T-23	16 PF	Parmia	11.61	4.26
T-24	16 PF	Premsia	7.97	2.54
T-25	16 PF	Protension	8.38	3.35
T-26	16 PF	Autia	11.07	3.27
T-27	16 PF	Shrewdness	11.42	2.72
T-28	16 PF	Guilt Proneness	10.62	3.68
T-29	16 PF	Radicalism	8.62	2.28
T-30	16 PF	Self-Sufficiency	8.96	3.10
T-31	16 PF	High-Self Sentiment	11.41	2.58
T-32	16 PF	High Ergic Tension	11.99	4.02
T-33	SIT	Social Judgment	19.24	3.63
T-34	SIT	Rec. Ment. State	23.70	6.74
T-35	SIT	Obser. Hum. Beh.	35.29	5.86
T-36	SIT	Mem. Names/Faces	9.47	6.37
T-37	SIT	Sense of Humor	10.83	4.18
T-38	W-G	Critical Thinking	51.12	10.23

TABLE 2
SUPERVISORS' TEST SCORES CORRELATED WITH MEASURES OF SUBORDINATES' MORALE AND PERFORMANCE AND SUPERIORS' RATINGS OF SUPERVISORS

Variable	M-1	M-2	M-3	M-4	P-1	P-2	P-3	P-4	P-5	P-6	R-1
T-1	02	18	30	14	12	34	27	14	06	07	09
T-2	06	—06	06	—07	09	01	07	02	00	05	20
T-3	15	—03	—02	—10	—08	—09	—20	11	—21	07	16
T-4	04	07	21	26	—04	21	10	17	03	08	02
T-5	06	—02	—04	—16	04	—08	—10	—09	—06	—04	—05
T-6	—06	—14	09	01	06	—09	—11	—18	12	—06	11
T-7	01	—07	—03	00	03	00	08	—10	—09	05	10
T-8	—28	—29	—03	—24	12	—10	—21	01	01	02	—06
T-9	—21	—26	—08	—12	05	—28	—44	—08	—02	09	00
T-10	—06	—17	—02	—12	—07	05	—11	12	13	—06	03
T-11	—10	—13	—03	—06	—06	—01	—15	04	05	04	11
T-12	—06	05	—05	03	—13	01	—13	04	—01	09	08
T-13	—02	—12	02	—04	05	03	03	14	—09	—08	06
T-14	04	07	10	17	—09	07	—08	07	01	05	26
T-15	—03	—06	—08	—02	07	20	—06	12	21	—27	25
T-16	02	—01	11	14	02	19	—04	23	11	05	—04
T-17	—13	—21	25	—04	—06	—24	—11	—18	—07	—03	03
T-18	21	09	23	14	10	12	18	—01	—01	08	11
T-19	10	—05	—12	—03	03	05	—14	05	—11	—10	04
T-20	—06	—09	04	—10	01	—07	—30	20	—02	—11	—11
T-21	05	01	17	01	10	—01	—18	14	10	03	—23
T-22	14	06	—15	03	00	—02	—12	—01	—03	14	01
T-23	—08	—09	—13	—09	03	—22	—31	—11	08	02	—15
T-24	—05	—05	22	14	—03	—08	21	—14	—04	—06	02
T-25	—16	08	12	01	06	00	20	—01	00	—05	—13
T-26	—13	—07	—05	—14	05	—19	—05	—12	—05	—01	—08
T-27	10	17	—25	06	—21	—03	08	01	08	00	—16
T-28	07	23	13	17	—07	05	06	21	—10	01	—23
T-29	—01	—15	—35	—09	—04	—17	—16	—07	14	01	02
T-30	—08	07	—10	09	07	24	25	08	05	11	00
T-31	—11	—20	—31	—34	09	—02	—09	—17	04	—08	06
T-32	00	12	14	10	16	20	27	16	—03	—02	—22
T-33	—06	19	25	05	03	17	25	03	—07	—11	—02
T-34	—04	13	23	10	02	20	22	03	01	11	00
T-35	—06	09	14	06	06	09	19	06	—09	—04	—16
T-36	10	06	—03	08	05	—02	06	—08	—22	09	12
T-37	00	21	17	06	12	30	25	16	02	05	05
T-38	—01	—02	12	05	14	04	13	—13	—12	—02	06

Note: For simplicity of presentation decimals are omitted. *r*'s of .24 or greater are significant at the .05 level. *r*'s of .31 or greater are significant at the .01 level.

each. Table 2 presents the correlations between the test variables and each of the criterion variables. Included are 27 statistically significant relationships involving 16 test variables and eight criterion measures.

D. DISCUSSION

The absence of spectacular relationships between any of the predictors and any of the criteria suggests strongly that caution must be exercised in the application of these kinds of standard tests to broad evaluations of individuals or groups in industry. Cursory examination of Table 2 shows that many of the seemingly relevant test variables showed only negligible relationships with measures intended to define some aspects of supervisory success. Since the criterion variable R-1 comes closest to the conventional criterion measure in this area, it is particularly instructive to trace the last column of Table 2. The many near-zero correlations are an indictment against a simple clinical interpretation of the scores.

The tests used in this study showed some useful and meaningful relationships with criterion measures which, if carefully organized and weighted, could provide quite satisfactory prediction of some of the criteria. For example, five test variables were significantly related to the M-3 criterion. These variables taken together produced a multiple correlation of .52 ($r = .46$ after correction for shrinkage) against that criterion. Also, five test variables were significantly related to the P-2 criterion; these variables produced a multiple correlation of .46 ($r = .39$ after correction for shrinkage) against the P-2 criterion. And, as a final example, eight variables were significantly related to the P-3 criterion. They produced a multiple correlation of .58 ($r = .50$ after correction for shrinkage) with P-3. Since it is very likely that the predictive utility of any combination of test measures will vary from setting to setting and from one type of supervisor to another (4), these examples are intended only as illustrations and certainly not as suggestions of variables that would be universally useful in making such predictions.

The main thrust of the data presented supports the criticisms of the "patent medicine" approach to industrial selection and prediction and makes clearer the hazards involved in developing "clinical impressions" of potential supervisory success on the basis of standard tests often used for that purpose. At the same time, the varying patterns of relationship of the test variables with the various criteria suggest strongly that as we move away from thinking in terms of a single criterion of success and move further in the direction of working with broader, multifaceted conceptions of success, testing devices like those of the typical "appraisal battery" may, indeed, provide useful information.

E. Summary

Thirty-eight variables yielded by nine standard psychological tests were studied as potential predictors of 11 criteria of supervisory success. The test variables showed some statistically significant correlations with all but three of the criterion measures and offered the prospect of very satisfactory pre-

diction for three of the criteria: morale concerning supervision, average number of suggestions submitted, and average number of awards received for suggestions.

Despite moderate success in relating standard psychological tests to specific multiple criteria of supervisory performance, the data clearly demonstrated the danger involved in the usual approach to forming psychological appraisals from test results.

REFERENCES

1. BARRETT, R. S. Guide to using psychological tests. *Harv. Bus. Rev.*, 1963, **41**, 138-146.
2. DUNNETTE, M. D. A modified model for test validation and selection research. *J. Appl. Psychol.*, 1963, **47**, 317-323.
3. GUETZKOW, H., & FOREHAND, G. A. A research strategy for partial knowledge useful in the selection of executives. In R. Taguiri (Ed.), *Research Needs in Executive Selection*. Boston, Mass.: Harvard Grad. Sch. Bus. Admin., 1961.
4. HOBERT, R. Development of moderator variables to enhance the prediction of managerial effectiveness. *J. Appl. Psychol.*, 1967, **51**, 50-64.

Department of Psychology
Muskingum College
New Concord, Ohio 43762

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 41-53.

REVERSAL AND NONREVERSAL SHIFTS IN NEUROLOGICALLY HANDICAPPED CHILDREN*[1]

Department of Psychology, California State College, Los Angeles

ROBERT H. REIHER, JOSEPH G. PHELAN, AND VERNON L. KIKER, JR.

A. INTRODUCTION

Research in conceptualization problems indicates that an ontogenetic progression in concept achievement may exist through the comparison of kindergarten children (7) and normal human adults (5).

This study was conducted to determine the relationship between children with neurological impairment and adults with neurological insult on conceptualization tasks involving reversal and nonreversal shifts as in Phelan and Gustafson (9). It also seemed important to determine the relationship that existed between both children and adults with neurological insult and normal adults, as in Kendler and D'Amato (5), and slow learning young children (7).

Kendler (4, p. 44) defines a concept as "the acquisition or utilization of . . . a common response to dissimilar stimuli." Several theoretical models have played an important role in influencing experimental designs. One such model is that of an S-R viewpoint. Both Pavlov (8) and Spence (10) have expounded this approach with reference to conditional differentiation and discrimination learning. A single unit S-R theory assumes a direct association between physical stimuli and overt responses. Spence, utilizing this paradigm, limited his predictions to inarticulate organisms. Human beings, however, learn to make covert responses to physical stimuli. These responses produce stimuli that mediate between the external stimulus and the overt response.

Hull (2) was among the first to use the mediational construct. According to Hull, external stimulation can initiate within the organism an event called the "pure stimulus act," a response which functions to produce additional stimulation which serves as a cue for overt responses. The cue-producing function of the mediator can itself be learned. Thus, this notion allows for the entry of prior learning, stored in memory, into a new associational process.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 6, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This study could not have been completed without the help of Melba Miller, Superintendent of the Diagnostic School for Neurologically Handicapped Children, Los Angeles, California; and Edna Thompson, Educational Director.

Kendler (4) suggested that experimental conceptual behavior can be described from an S-R framework in two ways. One, which seems to fit the behavior of lower animals and young children, is to represent a common response to dissimilar stimuli in a single unit manner (S-R): i.e., with each overt stimulus being directly associated with an overt common response. For human adults, a mediation framework (S-r-s-R) is employed. This indicates that instances (S) of a conceptual grouping are transformed into an implicit representational response (r) that results in a primary cue (s) for an overt sorting response (R). The nature of this representational response (r) is that it typically represents an "abstraction" process on the part of the *S*. Therefore, within this theory, concept learning is the acquisition of a categorizing response to common mediators evoked by various otherwise unique stimulus patterns. Conceptual behavior is thus a special mediated chain involving multiple converging linkages between external stimuli and responses, through a discriminatory mediational process.

An experimental technique currently employed in the investigation of concept formation within the mediation theoretical structure is that of the Reversal Shift (RS). This paradigm is a meaningful representative of a transfer of training problem in a two choice situation in which a person's previously correct responses are no longer correct and his previously incorrect responses now are correct. Two basic types of shifts are used. For a reversal shift, the subject must learn to assign all formerly positive instances to the negative category and *vice versa*. The same stimulus dimension is relevant to both concepts, but the stimuli to category concepts are reversed. A nonreversal shift requires the subject to make these assignments on the basis of a different relevant dimension. Therefore, if attaining a concept is a process of associating physical stimulus attributed with overt category responses, it should follow that the reversal shift should be more difficult than the nonreversal shift. This follows from the fact that some proportion of the associations learned in the first concept is valid after the nonreversal shift. Reversal shifts, in contrast, require a change in all initially acquired associations. Thus, if the number of new associations to be formed (or old ones to be extinguished) is a measure of problem difficulty, reversal shifts should take longer to complete than nonreversal.

Experimental studies using reversal and nonreversal shifts (5) with college students have demonstrated a more easily attainable reversal shift. Kendler and D'Amato reasoned that "a reversal shift occurred at a more rapid rate than a nonreversal shift because at the completion of the learning of the first concept, the symbolic cues appropriate to the second concept were available to *S*s in the reversal shift groups" (5).

In a similar type of problem, rats find a nonreversal shift easier than a reversal shift (3). It is possible that a single unit S-R theory accurately represents the behavior of rats, while mediational S-R theory is required to explain concept learning of articulate humans.

In an attempt to locate a transitional point between single unit and mediational theory, Kendler and Kendler (7) performed a study with kindergarten children, which showed that these children, as a group, executed a reversal and nonreversal shift at approximately the same rate. When these kindergarten children were divided into fast and slow learners on the basis of their performances on the first problem (training discriminations), slow learners performed during the second problem (test discrimination) according to the single unit theory; like rats, they found a nonreversal shift easier. Fast learners, on the other hand, performed in accordance with the mediational theory; like college students, they found a reversal shift easier. These results were interpreted as demonstrating that these kindergarteners, taken as a group, were in the process of developing mediating responses relevant to this task, and that some were further along than others. "The suggestion is strong, then, that as you ascend the evolutionary scale, organisms *acquire a greater capacity to generate cues* that enable them to make rapid RSs" (6, p. 11). If the importance, then, lies in the fact that humans utilize mediated language responses, a fast reversal shift may indicate that the subject is using a common mediated response to classes of stimuli. A person responding to the stimuli in a single-unit fashion should characteristically show a slow RS. The fact that young children exhibit a slow RS can be used to indicate that they do not utilize a common response to stimuli.

Goldstein and Gelb (1) studied intellectual impairment of soldiers suffering traumatic lesions in the frontal areas. There was in particular a distinct loss of ability to think conceptually. In sorting tasks involving multicolored yarns, these subjects showed a restricted concrete level of performance. In particular, they would not generalize widely or shift from one basis of classification to another. Subsequent investigations have justified the conclusion that individuals with brain injuries, especially in the frontal areas, suffer a loss of abstracting ability.

Heinz Werner (11, p. 240) states that the gradual growth of a unitary intellectual function involves specific changes in the process of abstraction itself: "One who can shift his point of view in a purposeful grouping activity is no longer subject to the forces of sensory stimulation. He is able consciously to perceive that objects have different qualities, any one of which may be taken as the point of departure for an ordering process. In other words, this development indicates an immensely important step away from an abstraction closely allied to sensory organization and toward an abstraction guided by deliberately selected categories, such as color, shape, number, size, etc."

Werner points out pathological examples which he feels are concomitant with his developmental approach. Utilizing the work of Head, Goldstein, and Gelb with brain-injured patients, he points out the similarities in behavior in so-called abstraction experiments between the brain-injured adult and child.

Phelan and Gustafson (9) have conducted research in reversal and nonreversal shifts in acute brain injured with injury diffusely localized patients at Sepulveda, San Fernando, and Brentwood V. A. installations. For the acute cerebrovascular accident patients, there seems to be an impairment in mediational processes. In contrast to normals, a reversal shift did not occur at a more rapid rate than a nonreversal. It is in conjunction with these results that the current research with neurologically handicapped children from the Diagnostic School for Neurologically Handicapped Children, Los Angeles, has been conducted.

B. METHOD

1. *Subjects*

The *S*s were 42 students consisting of 10 females and 32 males ranging in age from six to 15. All *S*s have demonstrable neurological impairment of disparate origin. A total of 18 *S*s were eliminated: nine could not solve the first concept; four could not perform the first stage, second concept; one subject was color blind; and the remaining four were eliminated for the purpose of equating groups. If these subjects were included in our final results, the level of significance of the differences comparing the relative effectiveness of reversal and nonreversal *S*s would all be the same. *S*s are resident patients at the Diagnostic School for Neurologically Handicapped Children, Los Angeles.

2. *Material*

The N. Y. U. Card Sorting Test, Form 2, was utilized. This test is patterned after the Wisconsin Card Sorting Test I, differing mainly in that it permits the testing of more abstract and complex concepts. In this experiment two sample cards were used: a large orange diamond and a small dark gray ellipse with pointed ends—similar in shape to a football. The form and color of the stimulus cards differed from those used in the response deck to avoid having any response card be a duplicate of either stimulus card. The deck consisted of 32 response cards which varied in terms of four forms (circle, crescent, square, and a [-shaped figure), four "colors" (black, medium gray, yellow, and red), and two sizes. The designs appeared on three-inch square cards with the designs themselves being either 1 inch or 2 inches in height.

The two major concepts used in this experiment were shape (S) and color (C). The shape concept required rectilinear shapes (square and [-figure) to be sorted below the large orange diamond, and curvilinear shapes (circles and crescents) to be sorted below the small gray ellipse. The color concept required achromatic figures (gray and black) to be sorted below the small gray ellipse.

3. *Design*

The main purpose of this experiment was to compare the relative effectiveness of a reversal shift as compared with a nonreversal shift. This could be accomplished by the 2 × 2 factorial design in which half of the *S*s learned initially the shape concept and the other half learned the color concept, then half of the *S*s in each of these groups learned a reverse shape concept and the other half of each group learned the reverse color concept. Thus four groups were formed: Shape-Reverse Shape (S-RS); Shape-Reverse Color (S-RC); Color-Reverse Shape (C-RS); and Color-Reverse Color (C-RC).

In such a design as this, *S*s in the nonreversal groups (S-RC and C-RS) during the learning of the second concept would receive fortuitous partial reinforcement of sorting responses associated with their first concept. For example, placing a yellow crescent below the large orange diamond stimulus card would be a correct response for *S* in Group C-RS during the learning of the first concept. It would also be a correct response for the same *S* during the learning of the second concept. In fact, all rectilinear achromatic and curvilinear chromatic cards would provide partial reinforcement effects for the nonreversal *S*s during their learning of the second concept. Those cards were therefore eliminated during the *initial stage* of learning the second concept. This resulted in *S*s of both the reversal and nonreversal groups receiving 100 percent nonreinforcement of their *sorting responses* which had been correct for the first concept. The elimination of these aforementioned cards had also other consequences. Only rectilinear chromatic and curvilinear achromatic cards were left in the response deck during the first stage of the second concept. Since these cards resembled one of the two stimulus cards (orange diamond, gray ellipse) *both in terms of shape and color,* they would be appropriately sorted below the same stimulus cards for both the reverse shape and

TABLE I

Schematic Representation of the Design of the Experiment

First Concept: Shape or Color	Second Concept	
	First Stage: Reverse Shape and Reverse Color	Second Stage: Reverse Shape or Reverse Color
Group S	Group S-R (SC)	Group S-RS
		Group S-RC
Group C	Group C-R (SC)	Group C-RS
		Group C-RC
	Group O-R (SC)	Group O-RS
		Group O-RC

reverse color concept. That is, during this first stage, the sorting responses of the response cards in the abridged deck would be *identical* for both reverse shape and reverse color concepts. This point is represented in Table 1 by the abbreviation R(SC) which indicates that in the first stage of the second concept both reverse shape and reverse color sorting responses are correct for all *S*s.

This condition necessitated the reinsertion of the discarded cards during the second stage of learning the second concept. In order to compare the relative effectiveness of a reversal and nonreversal shift, it was necessary to have response cards which required *different sorting responses* for the reverse color and reverse shape concept.

In order to discover whether the transfer effects resulting from a reversal or nonreversal shift were positive or negative, a control group (Group O) was added that did not receive any training on the first concept. The first exposure these *S*s had to the experimental routine involved the first stage of learning the second concept (O-SC). The pattern of reinforcement to which the control group was subjected was identical to that received by the two major experimental groups.

To summarize, the design, which is schematically represented in Table 1, was an extension of the 2 × 2 factorial design described at the introduction to this section. The extension was necessitated by the needs (*a*) to equate partial reinforcement effects for the reversal and nonreversal shifts and (*b*) to discover the nature (positive or negative) of the transfer effects produced by reversal and nonreversal shifts. The first was accomplished by using an abridged response card deck during the first stage of the learning of the second concept, which at one and the same time provided 100 percent nonreinforcement of the sorting response initially learned by the reversal and nonreversal groups and which also did not provide any differential reinforcement of the sorting responses associated with the learning of the reverse shape or reverse color concept. The second need was accomplished by adding a control group whose training began with the first stage of the second concept.

4. *Procedure*

The mechanics of the experimental procedure involved a display panel for each stimulus card below which a small cross appeared. At the start of the experiment the following instructions were orally presented by *E*:

> You will be shown a series of cards one at a time. Some of these cards will belong with this card (*E* points to the first stimulus card) and others with this one (*E* points to the second stimulus card). You must decide, for each of the cards that I shall show you, with which of these two cards (*E* points to the stimulus cards) it belongs. Indicate your choice by taking this pointer and gently tapping the mark which lies beneath the card of your choice. If your choice is correct I shall say "Right," but if it is incorrect I shall say "Wrong." Do you have any questions?

The transition between the various stages of the experiment was made without informing *S*s of the modification in either pattern of scoring or total number of response cards that were to be used. The *S*s in all the groups were required to reach a criterion of 15 successive correct responses in their learning of the first concept and during the second stage of the learning of the second concept. The criterion during the first stage of the second concept was 10 successive correct responses.

The *S*s had a limit of five decks (160 cards) to reach the criterion of learning the first concept and for the second stage of learning the second concept. An *S* who during the second stage of the second concept failed to solve the problem within the limit of five decks was arbitrarily assigned to a score of 160. When the 16-card deck was used during the first stage of the second concept, a maximum of six decks (96 cards) was permitted. The response deck during all stages of the experiment was presented in an order which was designed to preclude sequential learning.

C. RESULTS

The response measure used during all stages of the experiment was the number of trials, excluding the criterion trials to achieve the performance criterion. Table 2 reports the data associated with the learning of the first concept for normal adults, cerebrovascular accident *S*s, and neurologically handicapped children. The data are presented both in terms of the performance of the total number of *S*s in both Group C and S, as well as their respective subgroups. Table 2 includes a similar comparison of the normal adults, cerebrovascular brain-injured adults, and neurologically handicapped children. Results were highly skewed and, therefore, the nonparametric Wilcoxon's Matched-Pairs Signed-Ranks Test was used to evaluate the differences.[2] For the neurologically handicapped children, similar to the brain-injured adults, the differences between the total shape and total color groups were not significant. The differences between the two subgroups of neurologically handicapped children that were compared in the learning of the second concept (Group C-RC and C-RS and Groups S-RS and S-RC) were slight and insignificant (with the use of the authors' equating procedure). The total median scores for the neurologically handicapped in both the shape and color concepts, much like the brain-injured adult, suggest that unlike the normal adults these concepts do not follow the patterns of initial dominant responses to the experimental situation.

Table 3 presents the data for both areas of the first stage of the learning of the second concept. With the neurologically handicapped child, there were no significant differences between the reversal and nonreversal groups; however, there was a much higher range in the neurologically handicapped child's score, which is similar to those findings with the brain-damaged adult. Unlike the normal adults, the neurologically handicapped child in the O-RS and O-RC groups showed inconsistencies indicating a lack of clear-cut transfer

---

[2] Because of the fact that the results were highly skewed, the Wilcoxon Matched-Pairs Signed-Ranks Test was used for all comparisons.

Table 2

NUMBER OF TRIALS REQUIRED TO LEARN THE FIRST CONCEPT

Part A: Ss = Normal Human Adults*

Group	Mean	S D	Mdn	Range
S-RS	2.6	3.5	0	0-11
S-RC	2.4	5.6	0	0-23
Total S	2.5	4.7	0	
S-RS	11.3	28.1	3	0-116
S-RC	8.6	17.3	3	0-72
Total C	9.9	23.3	3	

Part B: Ss = Adult, Cerebro-Vascular Accident Cases - Diffuse Damage**

Group	Mean	S D	Mdn	Range
S-RS	15.58	14.19	15.5	8-25
S-RC	17.16	8.47	14.0	6-32
Total S	16.7	7.49	14.8	
C-RC	14.21	10.26	11.5	3-47
C-RS	18.63	7.80	14.0	6-43
Total C	16.21	10.77	13.1	

Part C: Ss = Neurologically Handicapped Children

Group	Mean	S D	Mdn	Range
S-RS	38.5	48.70	20.0	0-128
S-RC	31.5	39.94	11.0	0-106
Total S	35.0	43.02	15.5	
C-RC	32.42	28.91	24.0	4-72
C-RS	28.0	50.30	14.0	0-140
Total C	30.21	39.42	19.0	

*Normal Ss data reproduced from Kendler, H. H. and D'Amato, R. F., (5). Permission to reprint granted by Dr. H. H. Kendler, senior author, and by the American Psychological Association for The Journal of Experimental Psychology.

**Adults-Cerebro-vascular accident cases reproduced from Phelan, J. G., and Gustafson, C. W. (9).

effects from the first concept to the first stage of learning the second concept.

Table 4 compares the relative effectiveness of reversal over nonreversal shifts. This table compares the data of Kendler and D'Amato on normal adults, brain-damaged adults (Phelan and Gustafson) and neurologically handicapped children. These data indicate the effectiveness of learning during the second stage of the second concept. It appears that for the neurologically handicapped children in both the Groups S-RS and S-RC and Groups C-RS and C-RC, the reversal shift was not observed to occur at any significantly

Table 3

NUMBER OF TRIALS TO REACH CRITERION OF LEARNING
DURING THE FIRST STAGE OF THE SECOND CONCEPT

Ss = Normal Adults

Group	Mean	S D	Mdn	Range
S-RS	6.0	4.4	4	2-16
S-RC	4.1	4.0	3	1-18
C-RC	5.4	6.2	3	1-26
C-RS	4.9	3.5	4	1-16
O-RS	12.3	13.1	7	3-50
O-RC	18.0	24.1	7	2-91

Ss = Adults-Cerebro- Vascular Accident

Group	Mean	S D	Mdn	Range
S-RS	31.25	10.57	28	19-62
S-RC	35.47	17.02	26.5	20-73
C-RC	34.22	21.18	26	14-89
C-RS	35.23	15.52	31	18-63
O-RS	30.41	18.22	24	13-83
O-RC	28.90	16.39	24	10-76

Ss = Neurologically Handicapped Children

Group	Mean	S D	Mdn	Range
S-RS	35.62	27.39	25.5	2-80
S-RC	33.13	27.15	44.0	2-74
C-RC	23.0	18.65	30.0	3-50
C-RS	21.14	20.98	21.0	1-57
O-RS	31.83	21.21	29.5	2-62
O-RC	37.67	34.19	27.0	5-84

faster rate than the nonreversal shift. These results are consistent with those of Phelan and Gustafson in testing brain-injured adults.

In comparing the two control groups of neurologically handicapped children with their matched experimental groups, no significant information is found as to whether reversal or nonreversal shifts resulted in positive or negative transfer effects. These results, again, are similar to those of the brain-injured adult.

Table 4

NUMBER OF TRIALS TO REACH CRITERION OF LEARNING DURING THE SECOND STAGE OF THE SECOND CONCEPT

Ss = Normal Adults

Group	Mean	S D	Mdn	Range	p
S-RS	6.7	25.6	0	0-106	< .01
S-RC	50.4	57.7	20	0-160	
C-RC	3.1	6.5	0	0-33	< .01
C-RS	60.3	52.3	40	0-160	
O-RS	16.2	39.3	0	0-160	
O-RC	29.4	45.5	6	0-16	

Ss = Adults -Cerebro-vascular Accident

Group	Mean	S D	Mdn	Range	p
S-RS	43.75	39.02	26	16-160	< not signif.
S-RC	52.89	46.37	37	13-160	
C-RC	32.57	27.04	21	12-106	< not signif.
C-RS	46.63	43.81	30.5	21-160	
O-RS	56.75	34.78	43.5	27-160	
O-RC	51.0	39.93	37	29-160	

Ss = Neurologically Handicapped Children

Group	Mean	S D	Mdn	Range	p
S-RS	81.38	76.07	83	1-160	< not signif.
S-RC	75.50	73.82	58.5	1-160	
C-RC	87.43	69.72	60	14-160	< not signif.
C-RS	82.85	71.59	128	0-160	
O-RS	75.5	45.06	61	22-160	
O-RC	78.0	52.35	74	18-160	

The comparisons, then, between the control groups of O-RC and O-RS and the experimental groups S-RC and S-RS differ from those of the normal adult subject with reference to the theory that nonreversal shift produces a demonstrable transfer effect. These data, then, differ from those of the findings of Kendler and D'Amato but are consistent with those obtained from brain-damaged adults.

D. DISCUSSION

Kendler and D'Amato (5) concluded from their work on concept formation in normal adults that human concept behavior could be explained by an S-R formulation involving a mediational process. The results obtained from work with normal adults demonstrate a superiority of the reversal shift and suggest a positive transfer effect. According to the S-R mediational theory, upon completion of the learning of the first concept, the symbolic cues appropriate to the second concept would be present for the *S*s in the reversal group but merely connected to the "wrong" sorting response.

The neurologically handicapped child and quite possibly the brain-injured adult manifest consistently wide discrepancies between the results obtained from normal adults. The neurologically handicapped child seems to indicate an inconsistent functioning mediational process as defined in terms of the experimental operation. By comparison, many of the neurologically handicapped children took a large number of trials to achieve the first and second concept. Problems with a fixed response, memory impairment, and confusion were evident. Unlike the normal adult subjects, there seemed to be no marked transfer effects in either direction. This gives rise to an assumption that the possibility of an ineffectively functioning mediational process exists in the neurologically handicapped child.

E. SUMMARY

In the neurologically handicapped child, like the brain-injured adults in the card-sorting task, a facilitation of card sorting appears to occur even if the appropriate symbolic cues are connected to the wrong sorting response. The neurologically handicapped child, much like the brain-injured adult, does not seem to show a significant difference in the ability to effect a reversal shift more rapidly than a nonreversal shift. At the completion of the learning of the first concept, the symbolic cues for the second group do not seem to be any more present for *S*s in the reversal group than the nonreversal group, nor do they ever seem to be connected to the wrong response.

The results which Kendler predicted in reference to positive transfer effects were not found. Kendler interpreted his results to mean that the essential learning which is basic to performance in these tasks entails a complex association between external stimuli and overt category responses mediated by internal (and probably) verbal activity. Unlike simple conditioning, in which practice presumably increases the strength of S-RS connections, training in concept problems determines which verbal cues are dominant. If appropriate cues are strengthened during training, transfer will be facilitated. Results

with the neurologically handicapped child and the brain-damaged adult indicate that there was no significant transfer effect in either direction. It appears, then, that there may be an impairment in the same mediational process which facilitates the reversal shift in the normal adult.

REFERENCES

1. GELB, A., & GOLDSTEIN, K. Verber Farbennahmenamnesie nebst bemerkingen ueber das amnestischen ueberhaupt und die beziehung zwischen sprache und dem verhaltern zur umwelt. *Psychol. Forsch.*, 1925, **66**, 127-186.
2. HULL, C. L. Knowledge and purpose as habit mechanisms. *Psychol. Rev.*, 1930, **37**, 511-525.
3. KELLEHER, R. J. Discrimination learning as a function of reversal and nonreversal shifts. *J. Exper. Psychol.*, 1956, **51**, 379-384.
4. KENDLER, H. H. Concept Formation. *Annual Rev. Psychol.*, 1961, **12**, 447-472.
5. KENDLER, H. H., & D'AMATO, R. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. Exper. Psychol.*, 1956, **49**, 165-174.
6. KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal process in problem solving. *Psychol. Rev.*, 1962, **69**, 1-16.
7. KENDLER, T. S., & KENDLER, H. H. Reversal and nonreversal shifts in kindergarten children. *J. Exper. Psychol.*, 1959, **58**, 56-60.
8. PAVLOV, J. P. Conditioned Reflexes (trans. G. V. Anrep). London, England: Oxford Univ. Press, 1927.
9. PHELAN, J. G., & GUSTAFSON, C. W. Reversal and nonreversal shifts in acute brain injured with injury diffusely localized. *J. of Psychol.*, 1968, **70**, 249-258.
10. SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427-449.
11. WERNER, H. Comparative Psychology of Mental Development. New York: Sci. Edit, Inc., 1961.

Department of Psychology
California State College at Los Angeles
5151 State College Drive
Los Angeles, California 90032

Published as a separate and in *The Journal of Psychology,* 1969, **72**, 55-59.

THE STUDY OF SELF-DISCLOSURE IN EXPERIMENTAL INTERVIEWS*[1]

Veterans Administration Hospital, Salem, Virginia

FRED W. VONDRACEK

A. INTRODUCTION

A major aspect of interpersonal transactions is the continuous exchange of information. In special interpersonal transactions, such as counseling and psychotherapy, a large part of this information exchange concerns information about the self.

Jourard (3) has labeled this type of information exchange self-disclosure; and, subsequently, he developed the Self-Disclosure Inventory (SDI) for the empirical study of self-disclosure (5). The host of studies which followed the introduction of the SDI attests to the apparent significance of the field of self-disclosure. Only relatively recently, however, has the validity of the SDI been questioned. Significantly, none of the recent validity studies of the SDI were able to confirm its validity (1, 2, 7, 8).

In addition to this apparent lack of validity of the SDI, which has been the major research instrument in the field of self-disclosure for years, there has also been a conspicuous absence of experimental research in this area. It will, therefore, be interesting to develop new, preferably behavioral, measures of self-disclosure, and to attempt the application of appropriate experimental methods in the study of this important variable.

B. METHOD

1. *Subjects*

Sixty undergraduate males studying introductory psychology were recruited for an experiment in which they would be interviewed. The interviewers were five male and five female psychology graduate students. Each interviewer interviewed two *S*s each in a probing, a reflecting, and a revealing interviewing technique. The interviewing techniques were defined by the interviewer's verbalizations, and interviewers were trained in the use of these techniques prior to their first actual interview.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 6, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This article is based upon a dissertation submitted in partial fulfillment of requirements for the Ph.D. at the Pennsylvania State University. The author is grateful for the continuous help of the dissertation committee chairman, Richard Lundy, and the constructive support of the other committee members, Gary Galbraith, Leon Gorlow, and Hugh Urban.

2. *Procedure*

All interviews were transcribed on magnetic recording tape. After the initial introductory statement by the interviewer, which was identical in all three techniques, the interviewers observed a three-minute period during which they held their verbalizations to an absolute minimum. The *S*s' verbalizations during this period were used as base-rate data in the statistical analyses. After this three-minute period, depending on which technique the interviewer was using, a probing, a reflecting, or a revealing statement was made at specified intervals. If the *S*s did not terminate the interview after the initial 20 minutes, the interviewer was instructed to do so.

3. *Measurement of Self-Disclosure*

Two parameters of self-disclosure—amount and intimacy—were measured separately. The amount was determined by timing the *S*'s verbalizations during the interview, carefully excluding periods of silence and interviewer statements. Intimacy of self-disclosure was judged independently by two judges, in each case from the first two pages of the typewritten interview transcripts. The ratings were made on a Likert-type scale, representing a theoretical continuum ranging from a score of 1 (least disclosing) to a score of 7 (most disclosing). Both amount and intimacy were measured separately for the base-rate periods and the experimental periods. Transcripts were disguised so that the judges were unable to determine which interviewing technique had been used in any given interview. At the conclusion of each interview, the Jourard Self-Disclosure Inventory was administered by an assistant.

C. RESULTS

Product-moment correlations were calculated to assess interjudge reliabilities. Correlations were computed separately for the judgments of intimacy of the base-rate periods, the treatment periods, and the two periods combined. The coefficients ranged from .71 to .76.

A preliminary analysis of covariance of the amount of disclosure scores revealed that neither the individual interviewers, nor the interviewer's sex accounted for a substantial amount of the variance. It was found, however, that the interviewing techniques significantly affected the amount of self-disclosure produced by the *S*s (Table 1). A Tukey Gap Test revealed that the probing condition was most effective in eliciting amounts of self-disclosure, while no significant difference existed between the reflecting and the revealing techniques (Table 2).

TABLE 1
ANALYSIS OF COVARIANCE OF AMOUNT OF DISCLOSURE SCORES

Source of variance	*df*	Mean square	*F*
Techniques	2	158644.84	3.74*
Error	56	42387.39	
Total *df*	58		

* $p < .05$.

TABLE 2
MEANS AND STANDARD DEVIATIONS OF AMOUNT OF DISCLOSURE SCORES*

Measure	Probing	Reflecting	Revealing
$\overline{X}$	625.00	458.00	424.75
SD	182.04	228.72	274.52

Note: Tukey Gap Test: Critical value = 147.57, $p = .05$.
* The unit of measurement is seconds.

Analysis of covariance of the intimacy ratings indicated no differences in the intimacy of the disclosures made by any of the experimental groups. However, the intimacy ratings of the treatment periods were much higher (more intimate) than the intimacy ratings of the base-rate periods ($p < .001$).

Finally, the correlations between the measures of self-disclosure used in the present study and the SDI ranged from .10 to .20, and were not significant.

D. DISCUSSION

The experiment was designed to study the differential effectiveness of three interviewing techniques in eliciting self-disclosure. Furthermore, it was attempted to develop behavioral measures of self-disclosure.

One interesting finding suggested that a probing interviewer elicits significantly greater amounts of self-disclosure than a reflecting or a revealing interviewer. This apparently contradicts the findings of Jourard and Landsman (4) and Jourard and Richman (6), who reported evidence suggesting the operation of a dyadic effect in the disclosing behavior of individuals, in such a way that disclosure by one individual will in turn lead to reciprocal disclosure by the other member of the dyad.

Tuckman (9) has suggested a hypothesis which may account for the findings of the present study. He suggested that a probing technique, being more direct and instrumental than a revealing technique (or a reflecting technique), should be more effective in eliciting self-disclosure; since the revealing technique operates primarily as a stimulant to disclosure, and appears to be much more passive and indirect than the probing technique. It could be speculated, on the other hand, that these differences hold up only in initial or short-term contacts, and that in long-term interpersonal relationships the "dyadic effect" develops gradually.

The failure to discover systematic differences in the judged intimacy of the interviews appears also to be related to the relative briefness of the encounter between the *S*s and the interviewers. It appears that intimacy may increase as the duration of the encounter increases. The significant increase in rated intimacy between the base-rate periods and the experimental periods may partly reflect this increase in intimacy as a result of the increased duration of the relationship, although it probably also reflects the sudden increase in the interviewer's verbalizations at the conclusion of the base-rate period.

More generally, the present study demonstrates the feasibility of experimental research in the area of self-disclosure. Both measures of self-disclosure used in this study were found to have satisfactory reliabilities (10). Interestingly, their correlations with the SDI, a self-report measure, were rather low. This may be due partly to the fact that the SDI confounds the two parameters of self-disclosure—amount and intimacy—which were measured separately in the present study.

The preceding conclusions should not be interpreted to suggest that the SDI should be abandoned as a research instrument. However, it would seem advisable to examine further the relationship between the SDI and the behavioral measures of self-disclosure introduced in this study. Furthermore, the experimental study of self-disclosure would appear to be the tactic of choice in studying this much neglected aspect of interpersonal transactions.

E. SUMMARY

In order to test the differential effectiveness of three interviewing techniques in eliciting self-disclosure, 60 male *S*s were interviewed either by a probing, a reflecting, or a revealing interviewer. Self-disclosure was defined as consisting of two parameters—amount and intimacy—and measures were developed to assess them. Analysis of covariance suggested that the probing condition was most effective in eliciting high quantities (amounts) of self-disclosure, and that none of the interviewing techniques produced differences in the rated intimacy of the *S*s' disclosures. The importance of developing an experimental approach and valid and reliable measures in the area of self-disclosure was stressed.

REFERENCES

1. HIMELSTEIN, P., & KIMBROUGH, W. W. A study of self-disclosure in the classroom. *J. of Psychol.,* 1963, **55**, 437-440.
2. HIMELSTEIN, P., & LUBIN, B. Attempted validation of the self-disclosure inventory by the peer-nomination technique. *J. of Psychol.,* 1965, **61**, 13-16.
3. JOURARD, S. M. Personal Adjustment: An Approach Through the Study of Healthy Personality. New York: Macmillan, 1958.
4. JOURARD, S. M., & LANDSMAN, M. J. Cognition, cathexis, and the "dyadic effect" in men's self-disclosing behavior. *Merrill-Palmer Quart. Behav. & Devel.,* 1960, **6**, 178-186.
5. JOURARD, S. M., & LASAKOW, P. Some factors in self-disclosure. *J. Abn. & Soc. Psychol.,* 1958, **56**, 91-98.
6. JOURARD, S. M., & RICHMAN, P. Disclosure output and input in college students. *Merrill-Palmer Quart. Behav. & Devel.,* 1963, **9**, 141-143.
7. LUBIN, B., & HARRISON, R. L. Predicting small group behavior with the self-disclosure inventory. *Psychol. Rep.,* 1964, **15**, 77-78.
8. PEDERSON, D. M., & BREGLIO, V. J. The correlation of two self-disclosure inventories with actual self-disclosure: A validity study. *J. of Psychol.,* 1968, **68**, 291-298.
9. TUCKMAN, B. W. Interpersonal probing and revealing and systems of integrative complexity. *J. Personal. & Soc. Psychol.,* 1966, **3**, 655-664.
10. VONDRACEK, F. W. The manipulation of self-disclosure in an experimental interview situation. Unpublished Doctoral dissertation, The Pennsylvania State University, University Park, Pennsylvania, 1968.

Veterans Administration Hospital
Salem, Virginia 24153

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 61-64.

AN ANALYSIS OF THE INSTRUCTIONAL EFFECTIVENESS OF VISUAL ILLUSTRATIONS PRESENTED VIA TELEVISION*

University Division of Instructional Services, The Pennsylvania State University

Francis M. Dwyer, Jr.

A. Introduction

The use of varied visual illustrations to complement oral instruction presented via television is rapidly gaining wide acceptance at all levels of instruction. However, there is very little experimental evidence available as to how learners react to the various types of visual illustrations and to the variations in the amount and kind of realistic detail which they contain. In some cases it may be that the types of illustrations used to complement oral instruction may actually be hindering the learning process rather than facilitating it. Consequently, it seems necessary that guidelines be developed which will enable educators to select those types of visual illustrations for their televised presentations which will be most effective in promoting student achievement of desired educational objectives. A study by Dwyer (2) indicated that when students interacted with visually complemented programed frames for as long as they felt necessary to comprehend the information presented, different types of visual illustrations facilitated student achievement with different degrees of effectiveness.

The purposes of this study were (*a*) to measure the relative effectiveness of varied visual treatments possessing different amounts of realistic detail when the visuals are viewed for equal amounts of time, and (*b*) to determine which type of visual illustrations complementing oral instruction presented via television were most effective in facilitating student achievement on five criterion measures.

B. Method

1. *Subjects*

One hundred and sixty-two students in Speech 200 classes at The Pennsylvania State University were used as subjects.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 7, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

2. *Treatment Group*

Students in Group (Grp) I, the control group, received no illustrations of the heart, but viewed slides containing the names of the facts and processes of the heart as they were mentioned orally. Grp II viewed simple line representations of the heart. Grp III viewed detailed, shaded drawings. Grp IV viewed photographs of a heart model, and Grp V viewed realistic heart photographs.

3. *Procedure*

Students were placed in treatment groups by indicating which of the five testing sessions they would be able to attend; the experimental treatments were then assigned at random to the five groups. The audio presentation consisted of a 17-minute audio-taped presentation which described the heart, its parts, its internal functioning, and the simultaneous processes which occur during the systotic and diastolic phases. Each visual treatment contained 39 black and white slides which were synchronized with the audio-taped presentation, so that the slides presented visually relatively the same information that was being presented orally. Each instructional presentation was videotaped on an Ampex 660 B, and later presented to the respective treatment groups by means of the Telebeam Model A 912-A Television Projector which projected a 5 × 3-foot image. All groups received the same oral instruction, and viewed their respective type of visual illustrations for equal amounts of time.

4. *Criterion Measures*

Each student in each treatment group received a pretest, participated in his respective instructional presentation, and then received four individual criterion tests. Scores received on these tests were combined into a 78-item total criterion test. The objective of each test was as follows: (*a*) drawing test—to evaluate learning of specific locations of the various patterns, structures, and positions of the parts of the heart; (*b*) identification test—to measure transfer of learning: i.e., the ability to identify numbered parts on a diagram of the heart from information received in the instruction; (*c*) terminology test—to evaluate student knowledge of referents for specific symbols; (*d*) comprehension test—to measure understanding of the heart, its parts, and internal operations; and (*e*) total criterion test—to measure the student's total understanding of the concepts presented. The Kuder-Richardson Formula 21 reliability coefficients for the five criterion measures were: (*a*) drawing test .84; (*b*) identification test .81; (*c*) terminology test .81; (*d*) comprehension test .76; and (*e*) total criterion test .93.

C. RESULTS

The Hartley test for homogeneity of variance (6, pp. 94-95) performed on the pretest scores achieved on the Quick Scoring Mental Ability Test indicated that the treatment groups were drawn randomly from populations with common variance. The F-ratio resulted from the analysis of variance indicated that significant differences existed among the treatment groups on the drawing test ($F = 3.32$, $df = 4/161$, $p < .05$). No significant differences existed among the means of the remaining four criterion tests.

Comparisons among the individual means of the five treatment groups on the drawing test were conducted via Tukey's W Procedure (3, 4). The results indicated that the students viewing the abstract line presentation (Grp II) and the detailed drawing presentation (Grp III) achieved significantly higher scores than did those students receiving the oral presentation without visuals (Grp II > Grp I, $W = 2.80$, $n/v = 4/162$, $p < .05$; Grp III > Grp I, $W = 2.66$, $n/v = 4/162$, $p < .05$).

D. DISCUSSION

The results indicate that the use of visual illustrations to complement oral instruction is not an effective way to facilitate student achievement of all types of educational objectives. A number of possible explanations may be cited to explain the results obtained in this study: (*a*) since college students are generally selected from the upper two-thirds of the population in terms of verbal and conceptual ability, it seems that they are in a highly favorable position in terms of being able to learn from oral instruction; (*b*) the initial impact of the realistic detail in the visuals may have had the net effect of distracting the students from the essential learning cues in the oral and visual presentations; (*c*) student achievement of the types of learning objectives measured by the identification, terminology, comprehension, and total criterion test are not facilitated by visual cues similar to the ones used in this study; (*d*) since students in each treatment group viewed their respective televised presentation for equal amounts of time, those students who viewed the more realistic types of visuals may not have had sufficient time to study and comprehend adequately the additional information contained in the visual illustrations presented to them. The effectiveness of the line drawings seems to be in accordance with Attneave's research (1) which contends that only essentials travel through the nervous system to the brain and that lines bordering objects provide the essence of the information to be conveyed. Travers and his colleagues (5) have stated that visual information is stored in the nervous system in some form isomorphic with line drawings, and that this permits the individual to organize, remember, and reproduce edited information more easily than relatively realistic information.

E. Summary

A number of important generalizations emerge regarding the use of visual illustrations to complement oral instruction presented by means of television.

1. The use of visuals to illustrate subject content is not an effective means of improving achievement by college students of all types of educational objectives when students view their respective visuals for equal amounts of time.

2. The data indicate that the instructional effectiveness of visual illustrations is a function of the type of learning task they are designed to complement. This would seem to imply that each television instructor will have to define his objectives behaviorally and then select the appropriate visuals that have been found to promote most effectively student learnings of those specific objectives.

References

1. Attneave, F. Application of Information Theory to Psychology. New York: Holt, 1959.
2. Dwyer, F. The effectiveness of visual illustrations used to complement programed instruction. *J. of Psychol.*, 1968, **70**, 157-162.
3. Hopkins, K. D., & Chadbourn, R. A. A schema for proper utilization of multiple comparisons in research and a case study. *Amer. Educ. Res. J.*, 1967, **4**, 407-412.
4. Sparks, J. N. Expository notes on the problem of making multiple comparisons in a completely randomized design. *J. Exper. Educ.*, 1963, **4**, 344-349.
5. Travers, R. M. W., McCormick, M. C., Van Mondfrans, A. P., & Williams, F. E. Research and Theory Related to Audio-Visual Information Transmission. Salt Lake City, Utah: Bur. Educ. Res., Univ. Utah, 1964.
6. Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

University Division of Instructional Services
The Pennsylvania State University
103 Carnegie Building
University Park, Pennsylvania 16802

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 65-69.

FIELD INDEPENDENCE AND PERFORMANCE IN DISTRACTION* [1]

Department of Psychology, University of Texas at Austin

B. KENT HOUSTON

A. INTRODUCTION

While in the past few years an abundance of research has been done on the effects of stress on performance (8), comparatively little work has been done to determine what personality dimensions influence people's susceptibility to these effects of stress. After Witkin's (11) description of field dependence-independence was considered, it seemed that this might be one of those personality dimensions. The cognitive styles associated with field dependence-independence have been described as follows: "At one extreme (field dependence) there is a consistent tendency for experience to be global and diffuse; the organization of the field as a whole dictates the manner in which its parts are experienced. At the other extreme (field independence) there is a tendency for experience to be delineated and structured; parts of a field are experienced as discrete and the field as a whole organized" (11, p. 319).

It seems plausible then that field independent persons in a stressful situation might be better able than field dependent persons to separate and ignore the context of stressful stimuli and focus on task relevant cues. If this were true, field independent persons should perform better under stress than those who are field dependent. The experiment reported here represents an attempt to investigate this possibility.

B. METHOD

The general approach was that two groups of *S*s, a field independent (FI), and a field dependent (FD) group, performed three tasks first in quiet (Q) and later in a distraction condition (D). The two groups' performance on the three tasks in D (after adjusting for *S*s' performance in Q) was then evaluated.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 7, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This investigation was supported by United States Public Health Research Fellowship MH 15, 757. The author wishes to thank Eleanor Maccoby for her invaluable aid in completing this study and Ray Mulry for his constructive commentary on the manuscript.

Auditory distraction was chosen as a representative of many possible stressful situations (4). The three tasks were selected with the expectation that they would be sensitive to the effects of stress. They were Digit Span, Anagrams, and the Stroop Color-Word Interference Test. Digit Span (DS) was selected because it has been shown to be adversely affected by stress (10). The task was made additionally difficult by requiring *S*s to repeat the digits in the reverse order in which they had heard them. A 16-item Anagrams task (AGM) was chosen, since it is difficult and requires fairly constant attention. It was thought that these qualities would make the task vulnerable to the distracting effects of noise. The Stroop Color-Word Interference Test (CW) was used because it was thought that distraction would increase the interference involved and thereby have deleterious effect on *S*s' performance. In the CW task, four color names—red, green, blue, and orange—are repeated randomly on one sheet of paper for a total of 176 color names. Each color name was printed in one of the three colors other than the hue itself: e.g., the word "red" might be printed in blue, green, or orange ink. *S* was asked to name as rapidly as possible the color of ink in which each word was printed.

Scores on the Embedded Figures Test, EFT (6), an anxiety questionnaire (2), need achievement scale (3), and the American College Testing Program Examination (ACT) were obtained from 99 males and 95 females. On the basis of EFT scores, a field independent group (FI) and a field dependent group (FD) of *S*s were formed. The two groups were then equated on four variables that are possibly related to task performance in stressful situations: anxiety level, need achievement, intelligence (as measured by the ACT), and grade-point average. The final groups, each consisting of 20 females, performed the three tasks—DS, AGM, and CW—initially in Q. Three weeks later, *S*s again performed the three tasks, this time while wearing earphones[2] through which tape-recorded noises were played at a high but not painful intensity. *S*s were presented with two different forms of the tasks on the two occasions, the order of which was counterbalanced across *S*s. The order in which the three tasks were performed was also counterbalanced across *S*s.

Four scores were derived from the three tasks and constituted the dependent measures. The scores obtained were time and error scores for CW, number of correct digits recited backwards for DS, and the number of Anagrams correctly solved for AGM.

---

[2] Only one earphone was worn for the DS task in order to allow *S* to hear *E* present the digits.

C. RESULTS

The correlation between EFT scores and scores for anxiety, need achievement, and ACT for the entire sample of 99 males and 95 females are given in Table 1. The only correlation which was significant was that between EFT and ACT. The correlation between these variables was significant and similar in magnitude for both sexes.

TABLE 1
CORRELATIONS BETWEEN EFT SCORES AND SCORES FOR ANXIETY, NEED ACHIEVEMENT, AND ACT FOR MALES AND FEMALES

Sex	Anxiety	EFT and n Ach	ACT
Males	+.14	+.09	+.29**
Females	—.04	+.07	+.25*

* $p < .02$.
** $p < .01$.

There were no significant differences between the two forms of the three tasks for either of the Q or D conditions. The two forms of each task were therefore considered to be equivalent.

The performance of *S*s in FI was slightly superior to *S*s in FD on three out of the four measures—viz., DS, AGM, and CW time—but was slightly poorer on CW Errors. An analysis of covariance (9) on the task measures in D, adjusting for performance differences in Q, revealed that none of these group differences approached statistical significance.

D. DISCUSSION

That a correlation was obtained between EFT and intelligence as measured by ACT is consistent with Witkin, Dyk, Faterson, Goodenough, and Karp's (12) report that field independence correlates with intelligence. No correlation for either sex was found for need achievement or anxiety. The latter is in general accord with Bound (1) who reported that field dependence-independence was not related to measures of neuroticism.

The failure to obtain significant differences in task performance between FI and FD indicates that field independence is not linearly related to performance in distraction specifically and possibly performance in stress generally. This finding is in agreement with the work of Karp (7) and Holtzman and Bitterman (5). Karp (7) reports that field independence is not related to the capacity to ignore distracting stimuli in performing a task. The distracting stimuli Karp studied, however, were in the same—i.e., visual—

modality as that of the task relevant stimuli. The findings of this study indicate that Karp's conclusions may be extended in regard to the modality of the distracting stimuli. That is, field independence is not related in a linear fashion to the capacity to ignore distracting stimuli regardless of the modality in which they occur.

Holtzman and Bitterman (5) in a factor analysis of a wide range of measures relating to adjustment to stress included a vertical orientation task similar to one of Witkin's measures of field independence which correlates with EFT performance. The vertical orientation test did not load on the same factor as performance under stress.

One methodological criticism should be noted here. The effects of distraction and retesting are confounded in the D condition. A comparison group which performed the tasks in quiet the second, as well as the first, time would allow these effects to be separated. Since, however, no significant differences were obtained and the data are consonant with other findings, the lack of such a comparison group does not seem crucial.

E. Summary

It was hypothesized that field independent persons would perform better in a stressful situation than field dependent people. A group of 20 field independent female *S*s and a group of 20 field dependent female *S*s performed three tasks: Digit Span, Anagrams, and the Stroop Color-Word Interference Test, in both quiet (Q) and auditory distraction (D) conditions. Both had been equated on anxiety level, need achievement, intelligence test scores, and grade-point averages. An analysis of covariance for the scores on the three tasks obtained in D, adjusting for initial performance in Q, did not reveal any significant differences between the two groups. The findings of this study were considered as consonant with the reports of other investigators.

References

1. Bound, M. M. A study of the relationship between Witkin's indices of field dependency and Eysenck's indices of neuroticism. Unpublished Doctoral dissertation, Purdue University, Lafayette, Indiana, 1957.
2. Christie, R., & Budnitzky, S. A short forced-choice anxiety scale. *J. Consult. Psychol.*, 1957, **21**, 501.
3. Edwards, A. L. Edwards Personal Preference Schedule. New York: Psychological Corp., 1959.
4. Holtzman, W. H., & Bitterman, M. E. Psychiatric screening of flying personnel. VI. Anxiety and reactions to stress. *USAF, Sch. Aviat. Med. Proj. Rep.*, 1952, No. 21-37-002, Rep. No. 6, iii, 28 pp.
5. ———. A factorial study of adjustment to stress. *J. Abn. & Soc. Psychol.*, 1956, **52**, 179-185.

6. Jackson, D. N., Messick, S., & Meyers, C. T. Evaluation of group and individual forms of embedded-figures measures of field-independence. *Educ. & Psychol. Meas.*, 1964, **24**, 177-191.
7. Karp, S. A. Field dependence and overcoming embeddedness. *J. Consult. Psychol.*, 1963, **27**, 294-302.
8. Lazarus, R. S. Psychological Stress and the Coping Process. New York: McGraw-Hill, 1966.
9. McNemar, Q. Psychological Statistics. New York: Wiley, 1962.
10. Moldawsky, S., & Moldawsky, P. Digit span as an anxiety indicator. *J. Consult. Psychol.*, 1952, **16**, 115-118.
11. Witkin, H. A. Psychological differentiation and forms of pathology. *J. Abn. Psychol.*, 1965, **70**, 317-336.
12. Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R., & Karp, S. A. Psychological Differentiation. New York: Wiley, 1962.

Department of Psychology
Mezes Hall 211
The University of Texas
Austin, Texas 78712

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 71-83.

CONTRASTING PERSONALITY PATTERNS OF HIGH AND LOW ACHIEVERS AMONG COLLEGE STUDENTS OF ITALIAN AND IRISH DESCENT[*][1]

Department of Psychology, St. Peter's College; and Department of Psychiatry, New Jersey College of Medicine

ANGELO DANESINO AND WILLIAM A. LAYMAN

A. INTRODUCTION

Recent cross-cultural research points to significant differences in symptomatology and dynamics of abnormal behavior among various ethnic groups. In particular, patients of Italian and Irish descent have attracted the attention of many investigators (16, 18, 21). While germane to present cross-cultural interest, this study shifts its focus from the pathological to the normal pole of the behavioral spectrum. In addition, it was designed to test some of the theoretical assumptions and research results on psychocultural dimensions of achievement motivation presented by McClelland (13, 14, 15), Atkinson (3), Rosen (22), and Strodtbeck (27). Among the basic conclusions reached by these authors, the following are particularly pertinent to the present investigation: (*a*) certain factors exist in early family structure and subsequent cultural conditioning that lead to variations in achievement motivation; (*b*) these variations seem to persist during the process of assimilation to a new cultural environment; (*c*) achievement motivation is positively related to the degree of acculturation and to the rate of mobility of a specific ethnic group relative to others within the same culture; (*d*) in the American milieu high achievers tend to identify with the same sex parent.

The specific purpose of this study was, then, to investigate the effect of ethnic background on achievement motivation, parental identification, and religious ideology in male college students of Italian and Irish extraction, who were classified into High Academic Achievers (HAA) and Low Academic Achievers (LAA). It was hypothesized that, since high achievement motivation is positively related to the degree of acculturation, the difference between Italian and Irish High Academic Achievers on the variables mea-

---

[*] Received in the Editorial Office, Provincetown, Massachusetts, on February 13, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Suggestions and comments made by Dr. George W. Albee during the course of this study and by Dr. William E. Collins in the preparation of the manuscript are gratefully acknowledged.

sured should be less than that between the two Low Academic Achievers groups. Hence, the two HAA groups should reflect their original value system to a lesser extent than the LAAs. The rationale for selecting members of these two ethnic groups was that they present significant differences with respect to the specific psychocultural dimensions of achievement motivation as stated above. Each group differs in relation to the time of immigration into the United States and consequently to the length of acculturation to the American environment (1). There is some evidence that the mobility rate of the Italians is at present higher than that of the Irish (28). The original family structure and value system of the two groups, particularly the religious ideology, present contrasting patterns and characteristics.

B. METHOD

1. *The Sample*

From a total of 493 freshmen and sophomores in an all-male, Eastern Jesuit Liberal Arts College, 60 pairs of subjects were matched on the basis of intelligence, but differed in academic achievement. Each group of 60 High (HAA) and 60 Low (LAA) Achievers consisted of 30 students of Italian descent (Southern Italy) and 30 students of Irish descent, all comparable in socioeconomic background. All subjects were Catholics, of approximately the same age (18-20), and enrolled in the College of Arts and Sciences.

Students who reported at least one paternal and one maternal grandparent born in Southern Italy were considered Italians. Students with one or more paternal or maternal grandparent born in Ireland were considered Irish. Selection was made from freshmen and sophomore classes which were exposed to the same basic courses of instruction.

Upon entrance into the College, each student was given a college ability test (SCAT). SCAT scores for the 493 students ranged from 274 to 336, with a mean of 312 ($SD = 9$; median $= 312.4$). Selection of subjects was made from SCAT scores between 312 and 330. Differentiation of HAA and LAA subjects was made on the basis of actual performance through the first semester of college, since recent studies have indicated that the first semester General Point Average (GPA) is one of the most reliable predictors of academic performance in the four years of college (8). Sophomores whose academic achievement in the second and third trimester differed significantly from that of the first semester were eliminated from the sample. An index of achievement was established by comparing the percentile rank of each student's SCAT score with his percentile rank in academic grades during the first semester of college. HAAs showed academic achievement within the upper third of the class, and were performing by at least +10 above expectancy as measured by the index. LAAs showed achievement in the lower third of the class, and were performing at least —10 below expectancy, as measured by the index. On the basis of t tests that proved insignificant the HAA and LAA groups were adequately matched for intelligence, but were significantly different in academic achievement. Tests for homogeneity of variance among the groups yielded no significant results.

2. *Measuring Instruments and Testing Procedure*

The test battery was administered in a single afternoon and in the following sequence:

a. McClelland Achievement Motivation Test (14). Four pictures were used with standard instructions (14) for each subject tested under neutral conditions. They were: 1) a man working in a chemistry laboratory; 2) a boy with a vague operation scene in the background (TAT card 8 BM); 3) a man skiing down a steep hill; and 4) an original picture of a student in three-quarter profile, sitting at a desk reading a letter.

The scoring system used in this study was a slightly modified form of McClelland's *System C* (15), which yields a total need achievement score. Procedures and a detailed description of each scoring category are given by McClelland *et al.* (14, 15). A correlation of .89 was found for 45 randomly selected protocols scored by the authors and another qualified psychologist.

b. TAT Involvement Score. Degree of *parental involvement* was measured by stories written in response to two TAT cards (6 BM—older woman and man; 7 BM—older man and younger man). The two pictures were projected immediately after the McClelland test with standard TAT instructions and with a five-minute time limit for writing each story. Each story was scored for parental involvement on a six-point scale devised by Bieri, Lobeck, and Galinsky (5). Each subject was scored F (father) or M (mother) depending upon which story produced the highest involvement score. A score of zero was given if the two scores were equal. Quantitative scores were derived by assigning a value of +1 to the father and —1 to the mother. Of 50 stories independently scored by the authors and another qualified psychologist, there was 91 percent scoring agreement.

c. Ranking of Themes. Six indices pertaining to the scoring categories of the McClelland Achievement Motivation Test were chosen for particular analysis in this study. They were (*a*) G+ and G— (positive and negative affective states); (*b*) Bw and Bp (environmental and internal obstacles); and (*c*) SR and LR (short range goals and long range goals). The rationale for the selection of these variables was based upon the assumption that they represent contrasting personality patterns of Italian (G+; Bw; SR) and Irish (G—; Bp; LR) subjects (11).

Eight different themes selected from actual protocols of the McClelland Motivation Test and expressing all the combinations of the three pairs of variables taken one at a time were ranked by the subjects. The same four slides of the McClelland Motivation Test, as well as the same testing procedures, were used. Three minutes were allowed for the ranking of the eight themes for each slide. To obtain numerical scores, each subject arranged the themes on each slide from one to eight. Thus, by totalling the ranks of the four themes, concerning each of the six variables on each slide, a score was obtained for each subject for each variable on each slide. Each slide was scored separately. The total of the ranks for each subject was 36. For example, if the subject's score for variable G was less than the median of 18, he was rated a G+, if it was greater, he was rated a G—. This process was repeated for the B and R indices for each slide. Subjects who scored 18 for a given variable were eliminated.

d. Fichter Religious Scale. The Fichter Religious Scale (9), in a slightly modified form, was used to measure the extent to which the groups being tested agreed with the official position of the Catholic Church on certain moral, social, and religious questions. Five items taken from Schuyler's Religious Questionnaire (23) were added to Fichter's Scale for a total of 20 items; 10 of these were related to dogmatic and authoritative matters which require strict Catholic adherence, and 10 were nonauthoritative items which could be subjectively interpreted. Subjects answered each item with an *agree* (A) or *disagree* (D) response within a total time limit of four minutes. It was emphasized that the items merely represented a wide range of opinions. Each response indicating adherence to Church ideology was given a score of one; each nonadherence response was scored zero. The median for all HAAs on the authoritative scale was first determined. All Italian HAAs whose scores were below the median were classified as nonadherence; all whose scores were above the median were classified as adherent. The process was repeated for the Irish HAAs. Similarly, the median of the LAAs was determined on the authoritative scale and the Italian and Irish were classified as either *adherent* or *nonadherent*. Frequencies on the nonauthoritative scale were established in the same way.

e. Osgood's Semantic Differential. The Osgood Semantic Differential (20) was used with standard instructions and a four-minute time limit to measure the perceived similarity aspect of identification. A list of nine pairs of bipolar adjectives, to be rated on a seven-point scale, was employed. The subject rates himself, his father, and his mother on each of the nine dimensions. A high score for either parent indicates a perception of greater dissimilarity between the subject and the parent; a low score reflects a greater similarity. A perceived similarity with the father is scored by an F, with the mother by an M. By subtracting the M score from the F score, a difference score (D) is obtained. If the D is positive, the subject perceives himself as more like his mother; if negative, he sees himself as more like his father. If D is zero, the subject perceives himself to be equally similar to both parents (5). A quantitative score was derived by assigning a value of +1 to the father and —1 to the mother.

f. Similarity Perception Questionnaire. The Similarity Perception Questionnaire was used as another measure of parental identification. A special booklet was distributed to the subjects who were asked to respond to each question in as detailed a fashion as possible within a five-minute time limit. The content of the Questionnaire was suggested by the interview schedule used by the Funkenstein Biographical Questionnaire (10). In agreement with the procedure of Bieri *et al.* (5) the measure of *similarity perception* was based upon the answer given by each subject to the following question: *Which parent do you feel you are most like at the present?* If one parent was clearly indicated, a score of either F (father) or M (mother) was given; if both parents were mentioned, but one was given greater weight, a score of either F/M or M/F was given, and these were considered F and M responses respectively; if neither parent was relevant to the question, a score of zero was assigned. Quantitative scores were derived by assigning a value of +1 to the father and —1 to the mother.

The measure of *parental involvement* was based upon the answer given by each subject to the following question: *Which parent did you feel closer to as a child?* (4, 5). The question was interspersed with the other items of the Similarity Perception Questionnaire. The same scoring procedure outlined above was applied.

C. RESULTS AND DISCUSSION

1. *McClelland Achievement Motivation Test*

Results of the tests between the two ethnic groups of High Achievers (HAA) and of Low Achievers (LAA) are presented in Table 1. Italians had statistically higher scores ($p < .01$) than did Irish subjects in the HAA pairings. However, among the LAAs, Irish subjects had significantly higher

TABLE 1
TEST OF SIGNIFICANCE BETWEEN ITALIAN AND IRISH HIGH (HAA) AND LOW (LAA) ACADEMIC ACHIEVERS FOR MCCLELLAND MOTIVATION TEST SCORES

Group	Mean McClelland score	*SD*	*t*
Italian HAA	9.30	3.71	2.85**
Irish HAA	6.90	2.66	
Italian LAA	4.70	2.30	2.26*
Irish LAA	6.17	2.77	

* $p < .05$.
** $p < .01$.

scores ($p < .05$). A test of the difference between the differences of the Italian and Irish HAA and of the Italian and Irish LAA subjects yielded insignificant results. Thus, Italians tended to score at the extremes in motivation. Reports by McClelland (13), Rosen (22), and Strodtbeck (27) on the effect of ethnic and sociocultural factors on achievement motivation offer a possible interpretation for these findings. Italians, because of their more divergent ethnic background from the American culture and a greater language barrier, seem to have an initial disadvantage with respect to the Irish in the process of assimilation. Thus, Italians may more readily either accept or reject the challenge of their present sociocultural environment. The Italian HAAs, in order to overcome their greater initial handicap, must be more highly motivated than the Irish HAAs. Italian LAAs may not be able to free themselves so completely from previous sociocultural influences, which may expect modes of achievement different from those of the American environment. This different reaction of the Italian HAAs and the Italian LAAs could explain, at least partially, the differences in their achievement motivation level with respect to the two Irish groups.

Achievement motivation may also be related to the rate of status mobility in the two ethnic groups (6, 7). Although it has been found that Irish have a greater mobility than Italians (30), there are indications that the present Italian population has a relatively greater rate of acceleration with regard to status mobility than the Irish (28). Thus, the higher motivation of the HAA Italians could reflect this greater mobility rate, while the LAA Italians might be continuing at the status mobility level characteristic of early Italian immigrant groups. Concerning the often raised question as to whether mobility is a function of need achievement or *vice versa* (6, 7, 12), the analysis of the data point to a complex, multiple interaction including not only these two variables, but other psychocultural factors as well.

The intervals between the motivation mean scores of the LAA and the HAA groups were similar. Furthermore, the mean scores for all four groups were relatively low when compared to the maximum possible score. Further research should explore reasons for (*a*) the relatively large difference between the Italian and Irish HAA subjects, and (*b*) the general low motivation scores of this sample of Catholic college students which contrast with recent findings that indicate an increase in achievement motivation of Catholic as compared with non-Catholic groups (29).

2. *Ranking of Themes*

For each of the three sets of dichotomous variables, chi-square tests were used to evaluate the performance of the groups on each of the four slides (Table 2). For the G variable, the Italian HAAs had significantly greater G+ scores than did the Irish HAAs on slides 1, 3, and 4. Similarly, Italian LAAs produced greater G+ scores as compared with Irish LAAs on slides 1, 2, and 3. There were no significant differences between the *High* groups on slide 2, nor between the *Low* groups on slide 4. For the B variable, there were no significant differences between Italian and Irish HAA subjects on any slide. However, Italian LAAs showed a significantly greater concern than did Irish LAAs for external obstacles related to achievement goals on slides 1, 3, and 4. The R variable yielded no significant differences between Italian and Irish HAAs on any slide. However, Italian LAAs demonstrated a greater degree of short range (SR) quality in achievement goals, as compared with the Irish LAAs, on slides 1 and 3.

TABLE 2

Results of Chi-Square Tests of the Differences Between Italian and Irish High Academic Achievers (HAA) and Between Italian and Irish Low Academic Achievers (LAA) on G, B, and R Variables for Each Scene Used in the Ranking of Themes Test

	Scene			
Variable	Chemist	Student	Skier	Operation
G				
HAA	13.52***	.86	10.40**	11.52***
LAA	36.81***	14.16***	30.75***	1.01
B				
HAA	.52	1.04	.03	.51
LAA	34.83***	1.67	13.65***	12.54***
R				
HAA	3.59	.42	.88	2.93
LAA	12.03***	1.19	18.75***	.19

** $p < .01$.
*** $p < .001$.

The difference between the differences of Italian and Irish HAAs and of Italian and Irish LAAs on each variable for each slide was subjected to tests of significance (Table 3). For all three variables on slides 1 and 3, the differences between LAAs were significantly greater than the differences between HAAs. Slide 2 yielded no significant effects between those pairings. For slide 4, the difference between the HAA groups was significantly greater for the G— variable, the difference between the LAA groups was greater for the B— variable, and there were no significant effects found for the R— variable.

TABLE 3
RESULTS OF t TESTS OF THE DIFFERENCE BETWEEN DIFFERENCES FOR ITALIAN AND IRISH HIGH ACADEMIC ACHIEVERS (HAA) AND ITALIAN AND IRISH LOW ACADEMIC ACHIEVERS (LAA) ON G, B, AND R VARIABLES FOR EACH SCENE USED IN THE RANKING OF THEMES TEST

		Differences between means		
Variable	Scene	Italian and Irish HAA	Italian and Irish LAA	t
G	1	2.96	6.43	3.40***
	2	2.03	3.00	.65
	3	2.70	5.33	2.50*
	4	3.96	.73	2.23*
B	1	.33	5.93	5.49***
	2	.67	.77	.94
	3	.10	2.53	2.27*
	4	.24	3.80	2.59**
R	1	.66	3.73	3.23**
	2	.24	.97	.59
	3	.47	3.10	2.68**
	4	.87	.27	.09

* $p < .05$.
** $p < .01$.
*** $p < .001$.

These results may be due to (*a*) difference of the unique modal personality patterns in each ethnic group, and (*b*) higher degree of acculturation in the HAA group than in the LAA group. A possible explanation of the predominance of G+ for both Italian groups is that this index reflects the commonly observed ethnic characteristics of *extreme emotionality* and *life enjoyment,* as well as optimism and realism with respect to Italians (11). These traits are in opposition to the commonly observed Irish characteristics of *emotional inhibition, pessimism, idealism,* and *fantasy life* (24). The extreme poles (G+ for Italians and G— for Irish) for the affective state continuum appear to be the most distinguishing features between the ethnic groups. The above differences persist in the assimilation process and were even found in the psychopathology of patients of Italian and Irish descent (19, 26, 31). On the other hand, no significant difference between the HAA groups was found with respect to the B and R characteristics. It would seem that, with reference to external and internal barriers and to short and long range goals, the Italian and Irish HAAs have been assimilated to a point where there may no longer be any distinguishable difference relative to these traits. With LAA subjects, however, ethnic differences were found between the two groups with respect to all three variables. Thus, it seems that distinctive ethnic characteristics are more clearly delineated between Italian and Irish LAAs. On the other hand, the HAAs seem to be more acculturated to the American scene than the LAAs. This suggestion is based on the fact that differences between LAAs were greater than differences between HAAs with respect to all three variables.

3. *Fichter Religious Scale*

Chi-square tests were calculated for the high and low achievers with regard to adherence to the ideology of the Catholic Church with respect to dogmatic doctrine (*authoritative scale*) and nondogmatic doctrine (*nonauthoritative scale*). There were no significant differences between the tested groups (chi-square values ranged from .07 to 6.08). Further analyses by tests of the difference between the differences of the Italian and Irish HAAs and of the Italian and Irish LAAs also yielded negligible results. To account for these insignificant findings it would seem that differences in religious attitude related to ethnic backgrounds may have been overridden by the uniformity of the subjects' present religious education. These results are particularly interesting inasmuch as the original attitudes of the two groups were remarkably different (11).

4. *Parental Identification*

Two aspects of parental identification were considered: (*a*) *similarity perception* as measured by the Osgood Test and the Similarity Perception Questionnaire; and (*b*) *involvement* as measured by the two TAT cards and the Involvement Questionnaire. Subjects who expressed no similarity for either parent were eliminated from the analyses. Results (Table 4) indicate that Italian HAAs identified more closely with the father, while the Irish HAA tended to identify more closely with the mother, as measured by the *semantic differential* and *perceived similarity* scores. For LAAs the same *Italian father-Irish mother* relationship was found for the *semantic differential* score. However, the Similarity Perception Questionnaire, possibly because of its brevity and specificity, was not as general a discriminator. The Involve-

TABLE 4
RESULTS OF CHI-SQUARE TESTS OF THE DIFFERENCES BETWEEN ITALIAN AND IRISH HIGH ACADEMIC ACHIEVERS (HAA) AND BETWEEN ITALIAN AND IRISH LOW ACADEMIC ACHIEVERS (LAA) ON FOUR PARENTAL SIMILARITY PERCEPTION AND INVOLVEMENT SCORES

Groups	Perceived Similarity	Semantic Differential	Involvement Questionnaire	Involvement Score
HAA	24.36***	34.27***	.36	.24
LAA	.64	9.61**	.01	.58

** $p < .01$.
*** $p < .001$.

ment Questionnaire and the *involvement* score yielded no statistically reliable differences. Tests of the difference between the differences between Italian and Irish HAAs and Italian and Irish LAAs indicate that HAAs differed significantly more than LAAs on both the *perceived similarity* and *semantic differential* scores (Table 5). It would seem that the significant findings of the differences between both the HAA and LAA Italian and Irish groups reflect ethnic differences regardless of achievement motivation. However, since the difference between the Italian and Irish HAA groups is significantly greater than the differences between the Italian and Irish LAA groups, the present sample suggests that the more closely the individual identifies with the parent who plays a dominant role in his ethnic background, the greater is his achievement motivation (17). These results are of particular import in view of recent studies dealing with the relationship between achievement motivation and child-parent interaction and identification. In general HAAs tended to identify with the same-sex parent and seemed to have rather authoritarian mothers and less domineering fathers than did LAAs (2, 13, 25). It is interesting to observe that Italian HAAs in this sample identify with a paternal model which traditionally personifies authoritarianism, *bossism,* and

TABLE 5
RESULTS OF THE *t* TESTS OF THE DIFFERENCE BETWEEN DIFFERENCES FOR ITALIAN AND IRISH HIGH ACADEMIC ACHIEVERS (HAA) AND ITALIAN AND IRISH LOW ACADEMIC ACHIEVERS (LAA) ON PARENTAL SIMILARITY PERCEPTION AND INVOLVEMENT SCORES

Variable	Italian and Irish HAA	Italian and Irish LAA	t
Perceived Similarity	1.33	.28	3.38**
Semantic Differential	1.57	.85	2.50*
Involvement Questionnaire	.20	.13	.28
Involvement Score	.07	.03	.16

* $p < .05$.
** $p < .01$.

an unchallenged source of power. As a possible explanation for this finding one might suggest that the Italian HAAs, because of their more advanced acculturation into the American value system, are more tolerant of their original ethnic background and less susceptible to its negative influence than are Italian LAAs. Another interpretation may be that in the HAA group the father plays a less authoritarian and dominant role than in the original model. Consequently, the Italian HAAs might identify more easily with him along traditional patterns. In the case of the Italian LAAs, who are less assimilated than the Italian HAAs, paternal rigidity and authoritarianism seem to prevent the emergence of a strong achievement need in their sons. By contrast, the identification of the Irish HAAs with mother, which runs contrary to the same-sex parent model, could reflect a specific ethnic dimension, in that her traditionally more dominant family role as compared with the father could not only influence the sons' achievement motivation, but could also impinge upon their identification process. In the LAA groups a rather passive adherence to original ethnic values appears to influence negatively the performance of both Italian and Irish subjects. In the HAA groups, on the other hand, persistence of ethnic factors may positively stimulate the group members toward achievement and self-realization. Thus, in the pluralistic society of the United States, when group ties with the original ethnic culture become conflict-free in the process of assimilation, they may provide a greater impetus toward success and self-affirmation.

D. SUMMARY

This study investigated the effect of ethnic background on achievement motivation, religious ideology, and parental identification in a sample of 120 male college students, 60 of Italian and 60 of Irish descent. Each ethnic group included 30 High (HAA) and 30 Low (LAA) academic achievers. It was hypothesized that the difference between Italian and Irish HAAs on the variables measured should be less than that between the two LAA groups. Hence, the two HAA groups should reflect their original value system to a lesser extent than the LAAs. Data obtained from the test battery generally supported the hypothesis. However, contrary to expectation, Italian HAAs scored higher in the McClelland Motivation Test than their Irish counterpart. With regard to Parental Identification, the difference between the Italian and Irish HAAs was greater than that between the two LAA groups. As to the Fichter Religious Scale, no significant difference was found between groups. The relevance of psychocultural factors leading to high achievement was discussed.

REFERENCES

1. ARENSBERG, C. M., & KIMBALL, S. T. Culture and Community. New York: Harcourt, Brace & World, 1965.
2. ARGYLE, M., & ROBERTSON, P. The origin of achievement motivation. *Brit. J Soc. & Clin. Psychol.*, 1962, **1**, 107-120.
3. ATKINSON, J. W., *Ed.* Motives in Fantasy, Action, and Society. Princeton, N.J.: Van Nostrand, 1958.
4. BIERI, J., & LOBECK, R. Acceptance of authority and parental identification. *J. Personal.*, 1959, **27**, 74-86.
5. BIERI, J., LOBECK, R., & GALINSKY, D. M. Comparison of direct, indirect and fantasy measures of identification. *J. Abn. & Soc. Psychol.*, 1959, **58**, 253-258.
6. CROCKETT, H. J. The achievement motive and differential occupational mobility in the U.S. *Amer. Sociolog. Rev.*, 1962, **27**, 191-204.
7. ———. Social class, education, and motive to achieve in differential occupational mobility. *Sociolog. Quart.*, 1964, **4**, 231-242.
8. FEDER, D. D. Intriguing problems of design in predicting college success. *Educ. & Psycholog. Meas.*, 1965, **25**, 29-38.
9. FICHTER, J. H. Southern Parish. I. Dynamics of a City Church. Chicago, Ill.: Univ. Chicago Press, 1951.
10. FUNKENSTEIN, D. H., KING, S. H., & DROLETTE, M. B. Master of Stress. Cambridge, Mass.: Harvard Univ. Press, 1957.
11. GANS, H. T. The Urban Villagers. New York: Free Press of Glencoe, 1962.
12. LITTIG, L. W., & YERECARIS, C. A. Achievement motivation and intergenerational occupational mobility. *J. Personal. & Soc. Psychol.*, 1965, **1**, 386-389.
13. MCCLELLAND, D. C. The Achieving Society. Princeton, N.J.: Van Nostrand, 1961.
14. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. The Achievement Motive. New York: Appleton-Century-Crofts, 1953.
15. ———. A scoring manual for the achievement motive. In J. W. Atkinson (Ed.), *Motives in Fantasy, Action and Society*. New York: Van Nostrand, 1958. Pp. 179-204.
16. MURPHY, H. D. M., WITTKOWER, B. D., & CHANCE, N. A. Cross-cultural inquiry into the symptomology of depression. *Internat. J. Psychiat.*, 1967, **3**, 6-15.
17. OPLER, M. K. Culture and Mental Health. New York: Macmillan, 1959. Pp. 436-437.
18. ———. Culture and Social Psychiatry (rev. ed.). New York: Atherton Press, 1966.
19. OPLER, M. K., & SINGER, J. L. Ethnic differences in behavior and psychopathology. *Internat. J. Soc. Psychiat.*, 1956, **2**, 11-22.
20. OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. The Measurement of Meaning. Urbana, Ill.: Univ. Illinois Press, 1957.
21. PEARLIN, L. J., & KOLIN, M. L. Social class, occupation, and parental values: A cross-national study. *Amer. Sociolog. Rev.*, 1960, **31**, 466-479.
22. ROSEN, B. C. Race, ethnicity and the achievement syndromes. *Amer. Psycholog. Rev.*, 1959, **24**, 47-60.
23. SCHUYLER, J. B. Religious questionnaire. New York: Fordham Univ., Sociolog. Res. Lab., 1955.
24. SHANNON, W. V. The American Irish (rev. ed.). New York: Macmillan, 1966.
25. SHAW, M. C., & WHITE, D. L. The relationship between child-parent identification and academic underachievement. *J. Clin. Psychol.*, 1965, **21**, 10-13.

26. SINGER, J. L., & OPLER, M. K. Contrasting pattern of fantasy and motility in Irish and Italian schizophrenics. *J. Abn. & Soc. Psychol.*, 1965, **53**, 42-47.

27. STRODTBECK, F. J. Jewish and Italian immigration and subsequent status mobility. In D. C. McClelland *et al.* (Eds.). *Talent and Society*. New York: Van Nostrand, 1958. Pp. 259-268.

28. VECOLI, R. J. The People of New Jersey. Princeton, N.J.: Van Nostrand, 1965.

29. VEROFF, J., FELD, S. C., & GURIN, G. Achievement motivation and religious background. *Amer. Sociolog. Rev.*, 1962, **27**, 205-217.

30. WARNER, W. L., & SROLE, L. The Social Systems of American Ethnic Groups (Vol. 3, Yankee City Series). New Haven, Conn.: Yale Univ. Press, 1945.

31. ZOLA, T. K. Culture and symptoms: An analysis of patients presenting complaints. *Amer. Sociolog. Rev.*, 1966, **31**, 615-630.

Department of Psychiatry
New Jersey College of Medicine
88 Ross Street
East Orange, New Jersey 07019

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 85-92.

EXPLORATIONS IN METHODOLOGY: THE SCIENTIFIC QUESTION AS A SCHEMAPIRIC DIMENSION*

The Ohio State University

SVEN LUNDSTEDT

A. INTRODUCTION

In a recent article (3, p. 894), S. S. Stevens reminds us eloquently of the "curious antagonism which has sometimes infected the relations between measurement and statistics." The "schemapiric principle" joins together the ". . . informal, analytic, schematic features of model-building" with the ". . . concrete, empirical, experiential observations by which we test the usefulness of a particular representation." Notwithstanding arguments against Stevens' philosophy of measurement, one can find much comfort in his useful dictum of scientific responsibility that "Schematics and empirics are both essential to science, and full understanding demands that we know which is which" (3, p. 856). Consequently, this essay is not a quarrel with Stevens' essential thesis. Its aim is to question only the *extent* of his argument and to propose that another critical schematic dimension is equally as important initially in scientific endeavor. This dimension concerns the root of inquiry itself, the scientific question.

B. NATURE OF QUESTIONS

In an earlier paper (2, p. 229) the present author argued that there are major considerations making the valid structure of a good question: "(*a*) Is it important enough to worry about? Would an accurate answer add substantively to that which is already known? (*b*) Was it closely reasoned? In other words, does it cut to the heart of an issue? (*c*) What about the historical background against which the question is posed? (*d*) Possibly the most important—Is the question both logical and fitted to accepted standards of language usage?" Attention to these prior questions can affect the character of the direction of a system of thought. Increased emphasis may well be placed upon questions initially in addition to other methods for these reasons.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 17, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

C. Questions and Hypotheses

The usual canons of scientific method traditionally distinguish between the alternative and the null hypothesis, a convention which does not invalidate the author's argument. Initially, the chief motive for any inquiry *is* a question, not a formal hypothesis which is simply an edited version of one's question. An hypothesis always originates as a question, and remains like a question even in its declarative form because of its tentativeness and its appeal for proof or disproof. The relationship between the two has probably always been confusing. Their differences are formal and not substantive. In any event, research practices are often "quick and dirty" in this connection. Owing to such confusion we often lose sight of the primary value and utility of the research question, which is to be found in its *quality and relevance* to some important problem. Accordingly, there are only formal grounds for hairsplitting about the apparent differences between questions and hypotheses. Ultimately, we may not require the null hypothesis as an extra step in the research process. It may become necessary only to state one's research questions in their proper form, and then to follow this initial step with an appropriate empirical and logical test of the answers to them, which may simply be a quantified form of yes, no, or maybe.

D. Other Criteria for Choosing Questions

The criteria for developing a question are both substantive and formal. A concern with substance is always reflected in the question: Is it important enough to worry about? A question is trivial when it is superficial and does not add something tangible to the solution of a problem. It should cut to the heart of an issue. But substance has always been notably affected by form, and form by substance. If the form of a question is poor, its substance may not become apparent. The formal qualities of a question are determined by how well it fits accepted standards of logic and language.

At least six forms of research question can be distinguished, the most fundamental of which are the simple affirmative and negative. They receive this rank because they contain the least amount of information of any of the six. "Is X true?" and "Is X not true?" illustrate this elementary form. We recognize in them the familiar statement of identity.

A third and a fourth form are the conjunctive and disjunctive, so called because they connect two or more affirmatives with the word "and" or separate them with the word "or." For example, "Is it true that individual A has quality X *and* is it true that individual A has quality Y?" are illustrations. Replacing *and* with *or* makes it a disjunctive form. Accordingly, association and exclusion constitute the new information within both.

A fifth and a sixth form are the conditional and biconditional questions, so named because they affirm that one prior condition is dependent upon another subsequent one, or that the two are interdependent with one another. We recognize this as the language framework of the cause and effect relationship between any two things. Contained in them are not only identity and association, but dependency and interdependency as well. Both are clearly an improvement. A conditional is of the form, "Is it true that *if* some individual possesses quality X *then* will it result in that individual possessing quality Y?" An answer yes to this question specifies a relation of dependency, in one direction only. If the biconditional form is used, then the question, "Is it true that *if* individual A possesses quality X *if and only if* it is true that the same individual possesses quality Y?" expresses a relation specifying interdependency. A familiar, informal way of expressing such relations is to inquire if X causes Y and Y causes X (in this latter case) or to inquire if X only causes Y in the former.

A common mistake is to confuse the form and intent of a conditional or biconditional question with the form and intent of a conjunctive one; for example, to assume a cause and effect relationship where only an association is evident. As a rule, we may only *infer* from an association between two events that a cause and effect relationship exists between them; it is a matter of circumstantial evidence only. As a useful research strategy, however, we can often make good educated guesses about cause and effect based on answers to a conjunctive question.

Since all six of these types of questions are essentially propositional forms, they can be subjected to formal truth tests (such as those used in the analysis of logical fallacies) in order to determine whether or not they constitute sense or nonsense.

E. Schemapirics and Questions

We may ask if "Schemapirics" and the rules for good questions are joined, then do we get a better scientific tool? The answer would seem to be yes. We may envision as our universe of discourse the intersection of the set of all relevant questions with the set of all possible scales of measurement. Table 1 summarizes choices resulting from the intersection of the subset containing the six question forms with the subset containing the five traditional scales discussed by Stevens. Each cell identifies by number an initial research strategy. If our questions are important because they shape the direction of our thought, and to this extent are like self-fulfilling prophecies, then we are ahead of the game by knowing how we plan to begin. Our ability to predict and control is enhanced.

In Table 1, strategy 4.5 in the upper right-hand cell would traditionally be the best one. Not only does it employ a question form that takes account of the mutual influence of the independent and dependent variable, but it uses the ordinal scale of measurement said to have the most information. If, as Stevens says (3, p. 850), "The permissible transformations defining a scale type are those that keep intact the empirical information depicted by the scale," then the present author would add that similar criteria can also be used to identify a good question. Likewise, it is evident that strategy 1.1 is a weaker form containing less information.

TABLE 1[a]
EXAMPLES OF RESEARCH STRATEGIES

Type of question	Type of scale			
	Nominal	Ordinal	Interval	Ratio
Biconditional	1.5	2.5	3.5	4.5
Conditional	1.4	2.4	3.4	4.4
Conjunctive	1.3	2.3	3.3	4.3
Disjunctive	1.2	2.2	3.2	4.2
Affirmative or negative	1.1	2.1	3.1	4.1

[a] The index number 1.1 above indicates a scale and question combination having the least information, while the index number 4.5 indicates that combination with the most information.

Between these two extremes lies the popular arena of debate as to those answers to our questions which constitute greater or less empirical truth. As a rule, the physical and biological sciences traditionally come out much better in this debate than do the social sciences, although often the comparison has been wrongly an invidious one favoring, unjustly, the so-called "hard" sciences. In its own way the intellectual challenge for the social sciences is apparently no less demanding and the original distinction has become questionable.

But, physics and chemistry can indeed lay claim to many research strategies which are genuinely more rigorous. Their greater control of the research environment undoubtedly accounts to some extent for their eloquent progress. Like psychology, biology often comes off less well, by comparison, by virtue of its greater dependency upon its subject matter.

Even worse off are those social sciences which investigate human social behavior. They cannot expect to lay claim to much more than the bottom and top left hand portions of the table which contains strategies 1.1 to 2.5. There are those who would argue that unless the social sciences begin to approach the level of phenomenological control typical of these other sciences, they cannot hope to use effectively all strategies.

Whether an interval scale, for example, can be used effectively to express continuous data from human populations is a question yet to be resolved. Maybe the term "interval" can only be an approximate concept at best. Thurstone and Chave (4) ran head-on into this problem when they tried (unsuccessfully) to find interval measures of people's attitudes. Their problem lay in controlling the variations in the judgments required of those who formed the panel of experts who were to select the intervals in their attitude scales. Corrections later applied to Thurstone's original data, reported by Hovland and Sherif (1), demonstrated that such interval scaling was indeed subject to major personal biases, and that, such scales were only rough approximations of a true interval scale. Efforts continue to be made to improve this kind of situation, and some recent measure of success is evident. But on the whole, those who work on problems in the social sciences are restricted at present with the ordinal scale as their main tool. Knowing this one has to make the best of it, one facet of which can be to improve the quality and power of one's initial research questions.

The ethical problems raised by experimental manipulation and control, and often required by the conditional and biconditional form of question, are real obstacles in the social sciences. If logic and scientific method form the only path we wish to take to increase our knowledge of mankind's social and personal life, then in these fields we may always be severely limited in the extent to which we can pursue such knowledge freely. Even now it is no longer possible to pursue knowledge for its own sake when it is evident that the human price is too high. Increasingly, this facet of social and psychological research is becoming a recognized boundary beyond which certain research methods are forbidden. The social and behavioral sciences do indeed have rather unique constraints placed upon them by the social and ethical contexts in which they are expected to perform.

F. Complexities and Further Problems

One cannot in this brief essay touch upon all of the complex problems which naturally will come up when this new approach is taken to its logical and empirical conclusions. The 20 options in Table 1 only serve to hint at the nature of the terrain ahead. As one contemplates the options, linguistic and logical analyses appear to make up only some of the problems. As a rule, we always have to deal with the issue of the ultimate functional value of an extensive methodological schema too. Method for its own sake is hardly worthwhile. Somewhere our sense of proportion must enter into the picture. We have von Neumann's warning to lean upon, which Stevens cites too: "After much 'abstract' inbreeding, a mathematical subject is in danger of degeneration. At the inception the style is usually classical; when it shows signs of becoming baroque, then the danger signal is up" (5, p. 196). This might be said of research method.

G. PLACE OF STATISTICS

Nothing has been said thus far about the other side of the "schemapiric" issue: namely, the choice of an appropriate statistic to go along with the form of a research question and scale. This is not a trivial choice. Among the difficulties arising from wrong choices of this kind are those of claiming too much from one's data and those of claiming too little.

Measures of association or correlation are clearly appropriate for the conjunctive form. In special cases where prior methodology is experimental and so stipulated beforehand, such measures may even be used with the conditional and the biconditional. But, when no visible prior evidence of interdependency exists between two variables, then the associative form of statistic has an entirely different meaning empirically—a subtle and important difference. Space does not allow further analysis of the fit between questions, scales, and statistics here; but one which this author plans to continue. Table 2 summarizes the main thrust of this argument by identifying some categories of statistical methods appropriate in the different strategies.

The terms location, dispersion, and association are those used by Stevens and they involve such statistical measures, for example, as the mode, median, and arithmetic mean (location), percentiles, and standard or average deviation (dispersion), and rank correlation and product-moment correlation (association, dependency, and interdependency in special cases). The term differentiation applies to any technique which seeks to analyze the apparent difference between two variables, such as a correlation coefficient viewed negatively. Obviously, there is interpretive flexibility possible in the different measures.

Strategies 1.5, 1.4, 1.3, 1.2, and 1.1 are evidently limited by the constraints imposed by the nominal scale. The range of the simple affirmative or negative question appears to govern part of the choice of statistical measures in strategies 1.1, 1.2, 1.3, and 1.4. It is only when we move into the higher level strategies that we begin to experience more options. For example, in strategies

TABLE 2[a]

EXAMPLES OF RESEARCH STRATEGIES AND POSSIBLE STATISTICAL MEASURES

Type of question	Type of scale			
	Nominal	Ordinal	Interval	Ratio
Biconditional	Location Dispersion Differentiation (1.5)	Location Dispersion Differentiation Association Dependency Interdependency (2.5)	Location Dispersion Differentiation Association Dependency Interdependency (3.5)	Location Dispersion Differentiation Association Dependency Interdependency (4.5)
Conditional	Location Dispersion Differentiation (1.4)	Location Dispersion Differentiation Association Dependency (2.4)	Location Dispersion Differentiation Association Dependency (3.4)	Location Dispersion Differentiation Association Dependency (4.4)
Conjunctive	Location Dispersion Differentiation (1.3)	Location Dispersion Differentiation Association (2.3)	Location Dispersion Differentiation Association (3.3)	Location Dispersion Differentiation Association (4.3)
Disjunctive	Location Dispersion Differentiation (1.2)	Location Dispersion Differentiation (2.2)	Location Dispersion Differentiation (3.2)	Location Dispersion Differentiation (4.2)
Affirmative or negative	Location Dispersion (1.1)	Location Dispersion (2.1)	Location Dispersion (3.1)	Location Dispersion (4.1)

[a] The index numbers—e.g., 4.5—above are those used in Table 1 and have the same meaning.

3.4, 3.5, 4.4, and 4.5 ideally several additional statistical techniques may become allowable because of the assumptions underlying the questions and scales. In the final analysis all choices become governed, in turn, by the quantity and quality of one's initial data, which also may, or may not, rule out use of a given strategy.

H. SUMMARY

This paper began with the proposition that another critical dimension above and beyond those of schematics and empirics in their traditional sense is necessary initially in scientific endeavor. This dimension, the root of inquiry itself, is the scientific question. Several considerations making the valid structure of a good question were discussed, including the distinction between questions and hypotheses. Problems surrounding choice of one's research question were said to be substantive and formal. The substantive aspect of a question is its ultimate value and relevance. The formal aspect is to be found in its linguistic and logical structure.

Six formal questions were discussed, each containing successively greater explanatory power and called the simple affirmative, negative, conjunctive, disjunctive, conditional, and biconditional. This was followed by an analysis of some of the relations between questions, schematics, and empirics, as the latter two are defined and discussed by S. S. Stevens. The point is made that taken together, the form of a question and a scale can be used as a criterion for one's strategy in research. A wedding of "schemapirics" and questions should result in a better scientific tool. Finally, the relations of questions, scales and statistics to one another are discussed briefly indicating some of the broad initial choices that are possible within each of 20 strategies for research. There is need for further clarification of the relationship between questions, empirics, and other schemata as essential tools for research.

REFERENCES

1. HOVLAND, C. I., & SHERIF, M. Judgemental phenomena and scales of attitudes measurement: Displacement in Thurstone scales. *J. Abn. & Soc. Psychol.*, 1959, **47**, 822-832.
2. LUNDSTEDT, S. A note on asking questions. *J. Gen. Psychol.*, 1968, **79**, 229-239.
3. STEVENS, S. S. Measurement, statistics, and the Schemapiric view. *Science*, 1968, **161** (3844), 849-856.
4. THURSTONE, L. L., & CHAVE, E. J. The Measurement of Attitudes. Chicago, Ill.: Univ. Chicago Press, 1929.
5. VON NEUMANN, J. The mathematician. In *The Works of the Mind*, R. B. Heywood, *Ed.* Chicago, Ill.: Univ. Chicago Press, 1947. Pp. 180-196.

The Ohio State University
Division of Research
1775 S. College Road
Columbus, Ohio 43210

Published as a separate and in *The Journal of Psychology,* 1969, **72**, 93-98.

A COMPARISON OF TECHNIQUES FOR THE SCORING OF AVOIDABLE ERRORS ON THE RAVEN PROGRESSIVE MATRICES*[1]

Central Islip State Hospital

JOHN F. FRACCHIA, DIANE FIORENTINO, CHARLES SHEPPARD, AND SIDNEY MERLIS

A. INTRODUCTION

The Raven Progressive Matrices (RPM) was designed to measure the ability to form comparisons and reason by analogy, thereby providing an index of intellectual functioning (6). Both clinical practice and psychopathological dynamics suggest a relationship between type or degree of mental illness and disturbances in abstract processes, such as concept formation, comparative reasoning, and analogic reasoning. The RPM, while not primarily designed to assess the ideational effects of mental illness upon complex reasoning ability and intelligence, has been utilized in that capacity. Research findings (2, 3, 8, 9) appear contradictory with respect to RPM efficacy in distinguishing pathological influences.

Sheppard (8, 9), using an avoidable error measure based upon item difficulties, found significant differences in the number of RPM reasoning errors made by narcotic addicts having various MMPI profile patterns. Kasper used deviations from expected set scores as an index and reported a failure "to demonstrate the efficiency of PM in discriminating the influence of pathology" (3, p. 24). Jurjevich (2) found no strong relationship between the incidence of avoidable errors and degree of neurotic or characterological aberration when a weighted error scoring procedure was employed.

The purpose of this study is to investigate the extent to which defining and scoring RPM avoidable errors are significant factors in determining ability to detect differences in the abstract reasoning performance of MMPI diagnosed pathological groups. It is felt that defining avoidable errors in terms of empirically determined item difficulty levels results in a measure sensitive to the ideational effects of different types of mental illness.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 21, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This study was supported in part by Public Health Service Grant No. MH-05096-06 from the National Institutes of Mental Health.

B. METHOD

Three techniques for scoring avoidable errors were applied to RPM answer sheets of 88 narcotic users. The first method was derived from Raven's (6) concept of expected set score based upon the total test score. Raven (6) provides a table with the total number of items a testee should correctly solve in each set for a given score on the test. Additionally, items in each set are organized in increasing order of difficulty. If the testee, given his total score on the test, missed easier items in a set while solving more difficult items or missed more items than were expected for each set; these misses were scored as avoidable errors. Frequency of avoidable errors was summed across the five sets to establish a total Raven avoidable error score (R-AES).

The second method was called judges' rating of difficulty levels (JRDL). This technique defined an avoidable error as a failure to solve an item whose difficulty level was within the testee's range of ability as measured by his performance. Trained judges, using a table of empirically determined item difficulty levels (7), examined each RPM answer sheet, estimated the lowest item difficulty level the testee consistently passed, and scored the easier items missed as avoidable errors. The frequency of avoidable errors was summed to define the JRDL-AES.

The third method represented an attempt to attain increased objectivity by eliminating the need for judges and to overcome the apparent lack of consistent ordering of increasing item difficulty on the RPM. A frequency distribution of item difficulty levels, with a class interval of .10, was constructed from data reported in an earlier study by Sheppard *et al.* (7). Within each band of difficulty levels, the number of items needed to pass 70 percent of the items was determined. A cutoff item-difficulty value, the lower limit of the item-difficulty interval containing the most difficult item meeting the 70 percent criterion, was found for each testee. Items of easier item difficulty, higher than the cutoff value were scored as avoidable errors. These were summed to define the objective rating of difficulty levels (ORDL-AES). Product-moment correlation coefficients and percent agreement were used to measure the degree of association and congruence between the three techniques.

The 88 subjects studied represented subsamples of narcotic users established on the bases of MMPI profile patterns. Rules for profile interpretation defined by Gilberstadt and Duker (1), Marks and Seeman (4), and Meehl and Dahlstrom (5) were followed in constructing the MMPI profile groups. Data in Table 1 summarize the characteristics of this sample with regard to age, education, duration of heroin use, RPM centile, and MMPI scale scores.

Analysis of variance indicated nine significant F ratios on the MMPI scales at the .05 level or less. Subsequent t tests defined the reliability of profile assignment. Four profile patterns were established on the basis of MMPI T score elevations exceeding two standard deviations on the depression (2), psychopathic deviate (4), psychasthenia (7), schizophrenia (8), and hypomania (9) scales.

TABLE 1

MEAN AGE, EDUCATION, YEARS USING HEROIN, RPM AND IPAT CENTILES, AND MMPI T SCORES, STANDARD DEVIATION, AND F VALUES FOR FOUR MMPI PROFILE TYPES

	42′		428′		987′		49′			
	$N = 20$		$N = 19$		$N = 15$		$N = 36$		F	
Variable	M	SD	M	SD	M	SD	M	SD	ratio	p
Age	29	6.69	29	5.58	30	5.72	31	7.21	.55	NS
Education	10	2.60	9	2.03	9	2.10	10	2.42	.26	NS
Heroin use	9	4.62	8	3.80	11	4.67	10	4.98	.90	NS
Raven centile	48	29.60	50	25.83	41	30.00	58	25.89	1.46	NS
IPAT	38	28.45	29	26.05	32	29.04	36	28.34	.38	NS
L	48	6.47	51	7.90	48	9.13	49	7.81	.72	NS
F	53	11.81	54	23.35	46	27.43	54	13.36	.68	NS
K	53	7.95	51	8.32	44	8.90	51	8.41	3.67	.05
Hs	55	5.48	55	6.82	52	12.86	52	8.05	1.11	NS
D	71	6.72	73	9.69	62	10.72	58	9.14	14.86	.001
Hy	58	7.06	57	5.31	52	7.18	52	7.69	4.37	.01
Pd	76	7.32	81	8.04	70	10.01	76	7.97	4.63	.01
Mf	58	7.56	59	9.99	63	8.24	56	10.28	2.11	NS
Pa	51	8.16	67	9.52	65	6.30	52	5.73	25.48	.001
Pt	59	7.21	61	8.08	72	12.56	54	6.89	16.40	.001
Sc	55	4.88	71	7.80	73	20.17	55	4.93	21.47	.001
Ma	58	8.19	64	6.30	79	10.20	71	5.95	27.23	.001
Si	52	6.77	56	6.72	56	8.22	48	5.61	9.14	.001

The 42′, 49′,[2] 428′, and 987′ profile patterns were defined as secondary sociopathic, primary sociopathic, paranoid, and mixed schizophrenic personality types. Data seen in Table 1 regarding mean scale scores for these profile groups were consistent with clinical expectancy.

Tests of significant t are reported for groups having common overlapping scales. The 42′ group scored significantly higher on the depression scale than the 49′ group who scored significantly higher on the hypomania scale ($t =$ 5.48 and 6.70 respectively). The 428′ group scored significantly higher on the schizophrenia scale than the 42′ group, thus maintaining separation between these profile patterns on this dimension ($t = 7.52$). The 987′ group differed significantly from the 428′ group in elevation on the psychasthenia ($t = 3.00$)

---

[2] The authors wish to thank C. H. Ammons, Ph.D., and R. B. Ammons, Ph.D., Editors of Psychological Reports and Southern Universities Press for their permission to reproduce these data.

and hypomania ($t = 5.11$) scales. The 987′ group differed from 49′ group patients in elevation on the psychasthenia ($t = 6.43$) and schizophrenia ($t = 4.91$) scales.

With the definition of the groups established, analysis of variance was applied in order to test the significance of differences between the methods of defining avoidable errors with regard to their sensitivity in detecting functional differences in the reasoning performance of the four MMPI groups.

C. RESULTS

Measures of the degree of similarity between the three scoring methods, based upon the correlation coefficients and percent of RPM's scored between a single avoidable error of one another by each technique, indicated that ORDL and JRDL were most alike. Product-moment coefficients for ORDL-JRDL AES were .75, while ORDL-R and JRDL-R AES correlated .52 and .72 respectively. Although the correlations for ORDL-JRDL and JRDL-R scoring were almost identical, percent agreement measures revealed that 73 percent of the answer sheets were scored within one avoidable error of another by the ORDL and JRDL methods, while the corresponding values for JRDL-R and ORDL-R were 48 percent and 45 percent. Tests of significance for differences between proportions showed greater agreement between JRDL-ORDL than between either difficulty level method and R-AES at the .01 level. These findings suggest that the two difficulty level methods define and generally score avoidable errors in the same way.

TABLE 2
AVOIDABLE ERROR MEANS, AND STANDARD DEVIATIONS FOR THREE SCORING METHODS AND FOUR PROFILE PATTERNS

	Profile Patterns							
	42′		987′		428′		49′	
Scoring method	$N = 20$		$N = 15$		$N = 19$		$N = 34$	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
R	4.95	2.18	4.53	1.67	5.89	2.49	4.74	2.24
JRDL	3.30	2.53	3.80	2.83	5.63	3.23	3.41	1.82
ORDL	2.95	3.13	3.73	2.69	5.95	4.00	3.12	1.98

Table 2 presents a summary of the avoidable error means and standard deviations for profile groups according to the scoring method used. Analysis of variance indicated that both profile groups ($F = 4.10$, $df = 3/84$, $p < .01$) and scoring method ($F = 10.21$, $df = 2/6$, $p < .01$) were significant.

Tests of significance t indicated R-AES scored the profile groups differently from the other methods. Significantly more avoidable errors were assigned to the primary and secondary sociopathic groups by this method than by either difficulty level method (42:R $>$ JRDL, $t = 2.16$, $p < .05$; R $>$ ORDL, $t = 2.29$, $p < .05$; 49:R $>$ JRDL, $t = 2.68$, $p < .02$; R $>$ ORDL, $t = 3.15$, $p < .01$). Differences in avoidable error means between profile patterns only showed significant variation when scored by JRDL or ORDL methods. Both found the paranoids made more avoidable errors than sociopaths (JRDL:428 $>$ 42, $t = 2.45$, $p < .02$; 428 $>$ 49, $t = 3.14$, $p < .01$; ORDL:428 $>$ 42, $t = 2.55$, $p < .02$; 428 $>$ 49, $t = 3.38$, $p < .01$). These data support the hypothesis that the manner of defining and the method of scoring avoidable errors influenced the number of avoidable errors assigned to subject groups and thus appear to be factors in the ability of this index to distinguish the ideational effects of different kinds of mental illness upon intellectual reasoning.

D. Discussion

The data presented suggest defining and scoring RPM avoidable errors in terms of empirically determined item-difficulty levels yielded a measure of complex-reasoning-ability impairment capable of discriminating between MMPI profile-pattern-defined patient groups. Two scoring methods defining an avoidable reasoning error as "a failure to solve an item whose difficulty level was within the testees range of ability, as determined by his performance" (8, p. 1043) were shown to be sensitive to differences in pathological type. Both scoring methods indicated significantly more avoidable errors were made by *S*s with paranoid profile patterns than by *S*s with either primary or secondary sociopathic patterns. A scoring method using expected set scores as the basic defining concept for avoidable errors, applied to the same groups, did not show statistically significant differences in the number of avoidable errors found in the groups.

When mental illness affects intellectual ability in such a way as to produce irregularities in consistency and accuracy of performance, appropriately defined and objectively scored RPM avoidable error indices appear to be capable of estimating or assessing the magnitude of impairment in different patient groups. The appropriate definition, it seems, treats any item missed whose item-difficulty value is greater than an item-difficulty value representing the most difficult items an individual is capable of consistently solving as an avoidable error. The various definitions and scoring procedures for avoidable errors employed by different investigators may be one of the reasons results of previous research in this area appear contradictory (2, 3, 8, 9). The finding that conclusions based upon the same set of data would differ if the data were scored according to somewhat different avoidable error rationales provides support for this belief.

E. SUMMARY

The Raven Progressive Matrices of 88 narcotic users, divided into four MMPI profile pattern groups, were scored for avoidable errors with the use of three scoring methods. Two of the methods defining avoidable errors in terms of empirically determined item-difficulty levels indicated significantly more reasoning errors were made by patients with paranoid profiles than by patients with either primary or secondary sociopathic patterns. The third method, based upon expected set scores, did not score significantly different numbers of avoidable errors in the four groups. The data were interpreted as indicating that the manner of defining and that of scoring RPM avoidable errors are important factors in determining sensitivity to reasoning errors related to pathological ideation. It was suggested that where mental illness affects consistency of performance, appropriately defined and objectively scored avoidable error measures can be used to assess the impairment of intellectual functioning.

REFERENCES

1. GILBERSTADT, H., & DUKER, J. A Handbook for Clinical and Actuarial MMPI Interpretation. Philadelphia, Pa.: Saunders, 1965.
2. JURJEVICH, R. M. Avoidable errors on the Raven Progressive Matrices and psychopathological indices. *Psychol. Rep.*, 1967, **21**, 364.
3. KASPER, S. Progressive Matrices (1938) and emotional disturbance. *J. Consult. Psychol.*, 1958, **22**, 24.
4. MARKS, P. A., & SEEMAN, W. The Actuarial Description of Abnormal Personality. Baltimore, Md.: Williams & Wilkins, 1963.
5. MEEHL, P. E., & DAHLSTROM, W. G. Objective configural rules for discriminating psychotic from neurotic MMPI profiles. *J. Consult. Psychol.*, 1960, **24**, 375-387.
6. RAVEN, J. C. Guide to the Standard Progressive Matrices A, B, C, D, and E. London, England: H. K. Lewis, 1960 (Distr. U.S. by Psychological Corp.).
7. SHEPPARD, C., FIORENTINO, D., COLLINS, L., & MERLIS, S. Raven Progressive Matrices (1938): Normative data on male narcotic addicts. *Psychol. Rep.*, 1968, **23**, 263-268.
8. ———. Performance errors on Raven Progressive Matrices (1938) between sociopathic and schizotypic personality types. *Psychol. Rep.*, 1968, **23**, 1043-1046.
9. ———. Further study of performance errors on Raven Progressive Matrices (1938). *J. of Psychol.*, 1969, **71**(1), 127-132.

Central Islip State Hospital
Central Islip, New York 11722

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 99-100.

LAYMEN'S IDENTIFICATION OF ITEM CONTENT ON THE EYSENCK PERSONALITY INVENTORY*[1]

Department of Psychology, University of New England, Australia

---

GORDON STANLEY

A. INTRODUCTION

Recent work on the Eysenck Personality Inventory (EPI) by Gorman (2) and Salas (3) has indicated that the scale scores of the EPI are susceptible to successful faking by psychologically unsophisticated *S*s. Wiggins (4) has suggested that such faking may be possible because the *S*s know which behaviors are appropriate or inappropriate for the person's reference group. Previous workers have demonstrated that the EPI scales are susceptible to motivational sets; the present study is an attempt to determine how successful psychologically unsophisticated *S*s would be at identifying the psychological significance of the EPI items.

B. METHOD

1. *Subjects*

The *S*s were 31 adults of both sexes attending their first introductory lecture in psychology at an adult education center in northern New South Wales, Australia. The town is in a rural setting and most of the students were employed in nonprofessional occupations. Only one of the class had been to university and none had formal training in psychology.

2. *Procedure*

The *S*s were tested in a group setting and asked to complete the EPI according to standard instructions. The *S*s were given code numbers and told that their individual responses would be treated confidentially. After all *S*s had completed the inventory, the tests were collected by *E* and a second copy given to each *S* with the instruction to classify items according to content into three categories corresponding to the three scales of the EPI. A very brief definition of the extraversion (E), neuroticism (N), and lie (L) scale categories was given on the basis of the description given by Eysenck and Eysenck (1).

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 25, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This study was supported by a research grant to the author from the University of New England, Armidale, N.S.W., Australia.

C. Results

The mean correct identification of E scale items was 16.48 ($\sigma = 4.00$), the mean N scale identification was 15.42 ($\sigma = 4.74$), and that for the L scale was 4.61 ($\sigma = 2.57$). These identifications represent 68.66, 64.25, and 51.22 percent respectively of the scale items.

Product-moment correlations between the *S*s own performance on the scales under standard instructions and the number of items correctly identified on each scale were computed for each scale. The respective correlations for the E, N, and L scales were .115, .059, and .023, suggesting no systematic relationship between the *S*s' number of responses on the scale and their number of correctly identified items.

D. Summary and Discussion

Thirty-one people attending an introductory psychology lecture at an adult education center were asked to complete an EPI under standard instructions and under instructions to classify the items according to content into the three scale categories. Over half the items in each scale were correctly identified. There was no significant correlation between the number of items identified and the number of items endorsed for each scale. However it should be noted that in the present case the standard completion of the inventory did not occur in a highly ego-involving situation. The fact that "lay psychologists" with no professional training are able to identify successfully the psychological significance of over half the items on the inventory suggests that the instrument has little value for serious classification purposes under conditions where *S*s may be motivated to distort their responses.

References

1. Eysenck, H. J., & Eysenck, S. G. B. Manual of the Eysenck Personality Inventory. London, England: Univ. London Press, 1964.
2. Gorman, B. S. Social desirability factors and the Eysenck Personality Inventory. *J. of Psychol.*, 1968, **69**, 75-83.
3. Salas, R. G. Fakeability of responses on the Eysenck Personality Inventory. *Austral. J. Psychol.*, 1968, **20**, 55-57.
4. Wiggins, J. S. Social desirability estimation and "faking good" well. *Educ. Psychol. Meas.*, 1966, **26**, 329-341.

Department of Psychology
University of New England
Armidale, N.S.W., 2351
Australia

Published as a separate and in *The Journal of Psychology,* 1969, **72,** 101-107.

THE FACTORIAL DETERMINATION OF TYPES OF CONFORMING INDIVIDUALS*

Department of Psychology, Brigham Young University

SAUNDRA F. ZIMMERMAN, DARHL M. PEDERSEN, AND KAY H. SMITH

A. INTRODUCTION

Few attempts have been made to determine factor analytically types of conformity (1, 5). One approach to isolating types of conforming individuals is the obverse factor analysis method in which individuals are correlated over a set of measures. The present research included measures of personality characteristics and conformity. Conformity was measured by both paper and pencil tests and a behavioral group pressure situation similar to Crutchfield's (4). An obverse factor analysis was performed to determine factors which represented types of individuals associated with either paper-pencil measures or a behavioral measure of conformity.

B. METHOD

Measures of 38 attributes were obtained from 119 students at Brigham Young University. The following measures were used to assess conformity: conformity and independence subscales of the Survey of Interpersonal Values (6); the Overall-Agreement Scale (OAS) developed by Couch and Keniston (3) which is related to the acquiescence response set; and group pressure yielding (GPY) score determined by the amount of conformity in the group pressure situation. The following personality measures were administered: Rokeach Dogmatism Scale (8), Barron-Welsh Art Scale (2), Personal Orientation Inventory (10) with the subscales of self-actualizing values, existentiality, feeling reactivity, spontaneity, self-regard, self-acceptance, synergy, nature of man as constructive, acceptance of aggression, capacity for intimate contact, inner *versus* other directedness, and time incompetence *versus* time competence; Survey of Interpersonal Values (SIV) with the subscales of support, conformity, recognition, independence, benevolence, and leadership; Forms B and F of the IPAT Eight Parallel Anxiety Battery (9); Personal Data Questionnaire which obtained age, sex, year in college, frequency of church attendance, number of brothers and sisters, number of older brothers

---

* Received in the Editorial Office, Provincetown, Massachusetts, on March 3, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

and sisters, father's age, amount of father's education, father's occupation, mother's age, amount of mother's education, and parents' income. Grade point averages and scores on the American College Test were also obtained.

A 119 × 119 matrix of correlations between *S*s was obtained from the raw score matrix. Factors were extracted with the use of the principal axis method and rotated with the use of the normal varimax criterion (7). A multiple regression analysis was completed for each factor with factor loadings larger than .30 utilized as the dependent variable in order to ascertain which of the 38 variables accounted for the variance within each factor. Those independent variables which accounted for at least one percent of the total variance were selected and used to interpret the factors.

C. RESULTS

Twenty-nine factors were generated to account for the total variance and then rotated. The first 16 factors which accounted for 79 percent of the variance were included in the multiple regression analysis. Factors I through XVI accounted for 11.5 percent, 8.7 percent, 8.0 percent, 8.1 percent, 6.2 percent, 4.9 percent, 4.6 percent, 4.2 percent, 3.9 percent, 3.5 percent, 3.2 percent, 3.0 percent, 2.8 percent, 2.6 percent, 2.4 percent, and 2.3 percent of the variance, respectively. Only 11 factors were related to either GPY score, SIV conformity, OAS, or SIV independence. Table 1 gives the significant correlations of the personality variables with the factor loadings for each of the 11 factors.

The group pressure yielding (GPY) score or the behavioral definition of conformity was selected by the multiple regression process as predictive of factor loadings for three factors—V, VIII, and X. Factor V on which 14 *S*s received factor loadings beyond ±.30 was a "socioeconomic factor," since the higher correlations were with parents' income, father's occupation, father's education, and mother's education. Individuals who belonged to this group tended to be nonyielders in the group pressure situation and to come from families of high socioeconomic status. Factor VIII (seven *S*s received factor loading scores beyond ±.30) appeared to represent "ego-defensive" conformity. Individuals included in this factor demonstrated high GPY scores; yet indicated that they needed independence. They also denied having a need for recognition, had a lack of sensitivity to their own needs and feelings, and had a lack of spontaneity. Factor X (eight *S*s were included here) appeared to represent the "dogmatic, authoritarian" conformist. Conformers of this group tended to be dogmatic, rigid, and close-minded. They had relatively little tolerance or understanding for others and tended to be primarily aware of their own inner needs. They were low in leadership, basically inner-directed, and had no particular motivation to change their perspective.

TABLE 1

SIGNIFICANT CORRELATIONS OF THE PERSONALITY VARIABLES WITH THE FACTOR LOADINGS FOR EACH FACTOR IN WHICH EITHER GPY SCORE, SIV CONFORMITY, OAS, OR SIV INDEPENDENCE WAS SELECTED BY MULTIPLE REGRESSION TECHNIQUE

Personality variable	*r*
Factor V (GPY score predictive of factor loadings)	
Rokeach Dogmatism Scale	.63
Barron-Welsh Art Scale	—.74
Parents' income	.81
Father's education	.66
Mother's education	.72
Father's occupation	—.74
SIV recognition	.67
SIV independence	—.75
SIV benevolence	—.64
Nature of man as constructive	.64
Factor VIII	
Form B2, Anxiety Test	.94
Form F, Anxiety Test	.83
Number of brothers and sisters	.90
Number of older brothers and sisters	.96
Father's age	.95
SIV recognition	.81
SIV independence	—.78
GPY scores	—.71
POI feeling reactivity	.76
POI spontaneity	.73
POI synergy	—.81
Factor X	
Rokeach Dogmatism Scale	.82
SIV benevolence	—.77
SIV leadership	—.81
GPY scores	.90
Other *vs.* inner directedness	—.76
POI existentiality	.73
POI feeling reactivity	.90
POI self-acceptance	.83
POI acceptance of aggression	.75
POI capacity for intimate contact	.76
Factor III (SIV conformity predictive of factor loadings)	
Rokeach Dogmatism Scale	—.63
Form B1, Anxiety Test	.56
Factor III (*continued*)	
Form B2, Anxiety Test	.57
Form F, Anxiety Test	.57
Barron-Welsh Art Scale	—.45
Church attendance	—.64
Mother's age	.49
Parents' income	.45
Father's education	.53
Mother's education	.53
Other *vs.* inner directedness	—.84
Time competence	.45
Other directedness	—.82
Inner directedness	.85
POI existentiality	.87
POI feeling reactivity	.51
POI self-acceptance	.87
Nature of man as constructive	.49
POI synergy	.57
POI acceptance of aggression	.78
POI capacity for intimate contact	.62
Factor IV	
Rokeach Dogmatism Scale	—.66
Age of subject	—.53
Year in college	—.58
Number of brothers and sisters	—.49
Number of older brothers and sisters	—.49
Father's age	—.73
Mother's age	—.50
Parents' income	—.72
Father's education	—.78
Mother's education	—.47
Father's occupation	.64
SIV conformity	.67
SIV recognition	.77
SIV independence	—.73
GPY score	.72
Time incompetence/Time competence	—.86
Time incompetence	—.78
Time competence	.85
Inner directedness	—.60

TABLE 1 (*continued*)

Personality variable	*r*
Factor IV (*continued*)	
Self-actualizing value	.50
POI existentiality	.49
POI spontaneity	.46
POI self-regard	.85
POI self-acceptance	.71
Nature of man as constructive	.46
POI synergy	.75
POI capacity for intimate contact	.47
Factor XI	
Rokeach Dogmatism Scale	.50
Form F, Anxiety Test	.46
Barron-Welsh Art Scale	—.52
Year in college	.66
Church attendance	—.49
Number of brothers and sisters	—.89
Father's age	.58
Mother's age	.62
Parents' income	.55
SIV support	—.46
SIV conformity	.43
SIV recognition	—.51
SIV benevolence	.56
SIV leadership	—.53
Grade point average	—.46
Time incompetence/Time competence	—.57
Other *vs.* inner directedness	—.68
Time competence	—.55
Other directedness	—.69
POI existentiality	.57
POI feeling reactivity	.72
POI spontaneity	.51
POI self-acceptance	.50
Nature of man as constructive	—.54
POI synergy	—.42
POI acceptance of aggression	.78
Factor XV	
Year in college	—.71
Church attendance	.77
Father's age	.71
Father's occupation	—.74
Factor XV (*continued*)	
SIV conformity	.88
Self-actualizing value	—.65
POI existentiality	.75
POI acceptance of aggression	—.68
Factor I (SIV independence predictive of factor loadings)	
Form B1, Anxiety Test	—.83
Form B2, Anxiety Test	—.82
Form F, Anxiety Test	—.76
Sex of subject	—.51
Year in college	.49
Mother's age	.39
Parents' income	.50
Father's occupation	—.36
SIV support	—.77
SIV recognition	—.64
SIV independence	.52
SIV leadership	.64
Other *vs.* inner directedness	—.51
Time incompetence	—.44
Time competence	.46
Other directedness	—.61
Inner directedness	.69
Self-actualizing value	.70
POI existentiality	.42
POI spontaneity	.46
POI self-regard	.64
POI self-acceptance	.52
POI synergy	.54
POI acceptance of aggression	.65
POI capacity for intimate contact	.78
Factor XIV	
Rokeach Dogmatism Scale	.61
Sex of subject	—.80
Year in college	—.81
Number of brothers and sisters	.58
Mother's age	.77
SIV independence	—.77
SIV leadership	.75
Self-actualizing value	—.77
POI feeling reactivity	—.75
Nature of man as constructive	—.65
POI synergy	—.75

TABLE 1 (*continued*)

Personality variable	*r*	Personality variable	*r*
Factor II (OAS predictive of factor loadings)		Factor IX	
OAS	.94	OAS	—.71
Rokeach Dogmatism Scale	.69	Rokeach Dogmatism Scale	.56
Form F, Anxiety Test	.58	Barron-Welsh Art Scale	—.54
Age of subject	.65	Sex of subject	.48
Sex of subject	—.86	Year in college	.56
Church attendance	—.60	Number of brothers and sisters	—.75
Number of brothers and sisters	.46	Number of older brothers and sisters	—.52
Parents' income	—.82	Parents' income	—.59
Father's education	—.70	SIV conformity	—.68
Father's occupation	.73	SIV benevolence	.48
SIV recognition	.66	GPY score	—.63
SIV independence	.66	Grade point average	.54
SIV benevolence	—.84	Time incompetence/ Time competence	—.63
Grade point average	—.61	Other *vs.* inner directedness	—.89
Other *vs.* inner directedness	—.75	Time competence	.72
Other directedness	—.58	Other directedness	—.88
Inner directedness	.84	Inner directedness	.85
Self-actualizing value	.77	POI existentiality	.63
POI existentiality	.59	POI feeling reactivity	.69
POI feeling reactivity	.87	POI spontaneity	.76
POI spontaneity	.53	POI self-regard	.44
Nature of man as constructive	.59	POI self-acceptance	.58
POI synergy	.48	POI acceptance of aggression	.64
POI acceptance of aggression	.80	POI capacity for intimate contact	.61
POI capacity for intimate contact	.72		

The paper-pencil measure of conformity, SIV conformity, has been defined by Gordon (6) as a measure of "doing what is socially correct, following regulations closely and doing what is considered as acceptable and proper." It was selected by the multiple regression as accounting for some of the variance of four factors—III, IV, XI, and XV. Individuals ($N = 17$) grouped in Factor III were inner-directed, rather flexible, and acceptant of their feelings. It was a difficult factor to interpret, since SIV conformity was correlated very low with the factor loadings. Factor IV seemed to represent individuals ($N = 14$) who had high SIV and group pressure conformity. They were relatively dependent, seekers of recognition, and socially oriented. They appeared to accept both their strengths and weaknesses. They were from lower socioeconomic levels, had fewer brothers and sisters, and tended to be younger than the others in the sample. Factor IX contained individuals ($N = 14$) for whom the relationship between SIV conformity and factor loadings was moderate and positive. They were inner-directed, somewhat older, firstborn or only children, and expressive of their feelings. Factor XV contained individuals ($N = 7$) who demonstrated high conformity to societal demands. They came from families from a high socioeconomic level, had high church attendance, and appeared to have good social judgment in the application of values and principles.

There were two factors in which SIV Independence was predictive of factor loadings—I and XIV. Factor I contained individuals ($N = 26$) with high need for independence, low anxiety, and low need for support or recognition from others. On the other hand, they valued leadership highly and tended to have high self-regard. Factor XIV contained individuals ($N = 10$) who claimed to have a low need for independence and a high need for leadership. They tended to be close-minded, nonflexible, opinionated, and inhibited. Also, most of this group were females.

Overall Agreement Scale scores played a major role in two factors—II and IX. In the group which Factor II characterized were individuals ($N = 16$), primarily females, who, though they tended to score highly on "tendency to agree" (OAS), professed a need for independence. They had a tendency to be close-minded and conservative. On the other hand, they appeared to have a high capacity for making intimate, meaningful relationships and for expressing their feelings. Individuals ($N = 13$) grouped in Factor IX had low OAS scores (i.e., low tendency to agree) and high ability to withstand group pressure. They appeared to be inner-directed, to assess reality realistically, and to be spontaneous and expressive of their feelings. They also tended to have fewer brothers and sisters.

The results indicate that conformity is complex and multifaceted. Apparently there is no typical, conforming personality *per se*. Further research should ascertain how each of these types of individuals would behave across a variety of situations requiring varying degrees of conforming behavior.

D. SUMMARY

Four conformity measures (Survey of Interpersonal Values: Conformity and Independence subscales, Overall-Agreement Scale, and a yielding score for a Crutchfield-type group pressure situation) and 34 other personality measures were administered to 119 *S*s. An obverse factor analysis was completed with the use of the principal axis method with a normal varimax rotation. Sixteen factors which accounted for 79 percent of the variance were retained. Types of conforming individuals represented by the factors were

identified by performing a multiple regression analysis for each factor and by using the Ss' factor loadings as the dependent variable and the 38 measures as independent variables. Eleven factors which were related to at least one of the conformity measures were interpreted.

REFERENCES

1. BAROCAS, R., & GORLOW, L. Self-report personality measurement and conformity behavior. *J. Soc. Psychol.,* 1967, **71**, 227-234.
2. BARRON, F., & WELSH, G. W. Artistic perception as a possible factor in personality style: Its measurement by a figure preference test. *J. of Psychol.,* 1952, **33**, 199-203.
3. COUCH, A., & KENISTON, K. Yeasayers and naysayers. *J. Abn. & Soc. Psychol.,* 1960, **60**, 151-174.
4. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychol.,* 1955, **10**, 191-198.
5. ENDLER, N. S. Conformity analyzed and related to personality. *J. Soc. Psychol.,* 1961, **53**, 271-283.
6. GORDON, L. V. SRA Manual for Survey of Interpersonal Values. Chicago, Ill.: Sci. Res. Assoc., 1960.
7. KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika,* 1958, **23**(3), 187-200.
8. ROKEACH, M. The Open and Closed Mind. New York: Basic Books, 1960.
9. SCHEIER, I. H., & CATTEL, R. B. Temporary Handbook for the IPAT Eight-Parallel-Form-Anxiety Battery. Champaign, Ill.: Inst. for Personal. & Ability Testing, 1960.
10. SHOSTROM, E. L. Personal Orientation Inventory: An Inventory for the Measurement of Self-Actualization. San Diego, Calif.: Educ. & Indus. Testing Serv., 1963.

Department of Psychology
Brigham Young University
Provo, Utah 84601

Published as a separate and in *The Journal of Psychology*, 1969, 72, 109-114.

BIOGRAPHICAL INVENTORIES AND STYLE OF LEADERSHIP*

Department of Psychology, University of Windsor, Canada

JOHN J. LA GAIPA[1]

A. INTRODUCTION

A common practice in developing biographical inventories is to select items that discriminate against a criterion with little regard given to why an item works or does not work. Owens (5) cautions that a blind empirical approach is less likely to result in an effective instrument than one based on a theoretical rationale. Williams (9) found that items derived from specific hypotheses were more discriminating than items not based on any rationale.

For many jobs, no rationale exists for the development of items. This is not so in the area of leadership. Though research has demonstrated the effects of leadership style on group productivity and morale, little of this has been applied to the selection of leaders. Fielder (1) questions the selection of leaders according to their preferred style of leadership. Fielder points out that there is no best style—it depends on the situation. Subgroup analysis, however, may be one solution to this problem.

The main interest in this study is the development of a biographical inventory derived from research indicating that the same style of leadership is differentially evaluated at different positions in a hierarchy. Moore (4) found in his factor analysis of ratings of Air Force NCOs that subordinates described NCOs in terms of human relations and superiors described NCOs in terms of their ability to "get the work done." The present author hypothesized that officer candidates who prefer task-oriented to human-relations-oriented NCOs are more likely to be successful leaders.

B. METHOD

1. *Subjects*

The *S*s were 493 junior officers at the U.S. Navy Officer Candidate School. These line officers were recent college graduates recruited from civilian sources.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on March 7, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] A portion of this paper was read before the Southeastern Psychological Association, Roanoke, Virginia, April, 1968. The opinions and conclusions contained herein are those of the author and are not to be construed as representing those of the Department of the Navy.

2. *Material*

a. Biographical Information Blank (*Army BIB*). This 99-item scale contains life history, demographic, and attitude items. The scale is empirically keyed.

b. Risk Scale. This 73-item scale is a revision of a BIB developed by Torrance and Ziller (7). A high score indicates an active and aggressive person who develops a male role early in life.

c. Cooperation Scale. This scale contains 28 forced-choice tetrads (3). The examinee is asked to select the two statements that are "most descriptive" and "second most descriptive" of himself.

d. Leadership Orientation Scale. This 56-item scale was derived from the Air Force study cited above on differential leadership expectations (4). Statements describing NCO characteristics valued by superiors were paired with descriptions of NCO characteristics valued by subordinates. The examinee is asked to select the statements which best describe the petty officer that he, as an officer, would prefer to have in charge of the enlisted men.

e. Peer Ratings. A measure of each officer's "aptitude for leadership and military service" was obtained during the 11th week at the Officer Candidate School. Each officer rates all the men in his section. Each man's peer rating is based on the ratings he received from about 30 men in his section.

f. Officer Fitness Report. This is the official Navy criterion measure. An overall index of performance for each officer was obtained by averaging his scores on the following graphic rating scales: Performance, Desirability, Professional Knowledge, Cooperation, Judgment, Promotion Potential, Leadership, and Overall Rating.

g. Junior Officer Activity Checklist (*JOAC*). This 20-item checklist was used to evaluate officers in the Fleet. This experimental checklist describes critical elements of job performance required of officers aboard destroyer-type vessels. Each definition of a critical requirement is followed by two illustrative specific incidents. The reporting officer indicated how often in the preceding six months he had observed the behavior described in the item. The areas of performance included responsibleness, discipline, coordination, consideration, communications, training, and specific job duties. A reliability estimate of .79 was obtained by the Spearman-Brown prophecy formula.

3. *Procedure*

The experimental biographical inventories and other psychological tests were administered to the officer candidates shortly after they entered the Officer Candidate School. The officers were followed up to their duty stations a year later at which time criterion data were collected. To increase homogeneity with regard to the job performance being evaluated, the data were analyzed separately for the 264 officers assigned to Fleet duty and for the 229 officers assigned to Shore duty.

Product moment correlations were computed between the scores on each biographical inventory and the two criterion measures. With the exception of the Army BIB, which predicted performance of officers in Shore billets ($r = .22$; $p < .01$) as measured by the Fitness Report, none of the experimental inventories yielded validities higher than .16. The four inventories were subjected to item analysis in the hope of raising the validity coefficients.

The officers in each group were randomly assigned to item analysis and cross-validation samples. The item analysis sample contained 129 officers in Shore billets and 164 officers in Fleet billets. The cross-validation samples contained 100 officers in Shore billets and 100 officers in Fleet billets. Items were selected that differentiated high from low criterions beyond the .20 level of confidence. It was found that very few items discriminated both in the Shore and Fleet samples. Therefore, separate scoring keys were developed for each sample. These new keys were used to rescore the answer sheets of officers in the cross-validation samples. Then, product moment correlations were again computed between the total scores yielded by the revised keys and the official criterion measure.

C. RESULTS

The validities of two of the four biographical inventories were not significant in the Shore sample. The validity of the Army BIB did not hold up under cross-validation ($r = .05$; $p > .05$). The validity of the Risk Scale did not increase as a result of selecting the most discriminating items ($r = .12$; $p > .05$).

The Leadership Orientation Scale correlated .21 ($p < .05$), and the Cooperation Scale correlated .30 ($p < .01$) with the Officer Fitness Report. The correlation between these two predictors was —.01. Adding the scores into one composite score yielded a correlation of .37 ($p < .01$) with the criterion. The practical value of this validity coefficient was evident when it was compared with the predictive value of other measures. With the exception of OCS Peer Ratings ($r = .26$; $p < .01$), nothing in the battery of experimental and traditional aptitude and achievement tests, college grades, training school grades, and interview ratings had any value for predicting officer success.

None of the biographical inventories predicted officer success in Fleet billets as measured by the Fitness Report and the JOAC. The Leadership Orientation Scale had the highest validity against the JOAC ($r = .16$; $p > .05$). Peer Ratings, however, predicted the Fitness Report ($r = .25$; $p < .05$) and the JOAC ($r = .29$; $p < .01$).

D. Discussion

The results support the hypothesis underlying the construction of the Leadership Orientation Scale for the Shore based officers. Preferred leadership style was associated with later success in a leadership role as measured by the official criterion measure. Officer candidates who preferred task-oriented leadership to human-relations-oriented leadership were themselves rated higher on the job as officers.

The finding that "human relations" officers do not necessarily perceive themselves as cooperative was rather surprising. Since both measures predicted the criterion, the combined measures yielded a higher level of prediction. The results suggest that an officer with a low self-image of cooperativeness who emphasizes "getting the job done" is less likely to receive as high a leadership rating as an officer who is both cooperative and task-oriented. Perhaps, group productivity was lower under the task-oriented, noncooperative officer. This possibility would be consonant with the finding by Katz, Maccoby, and Morse (2) that the production-centered supervisors achieved a lower level of production than employee-centered supervisors. The value of combining measures of leadership style and a personality dimension is also consistent with Pelz (6) who suggests that it is an oversimplification to try to relate employee-oriented *versus* task-oriented behavior to productivity. The writer suspects that task-oriented officers varying in cooperativeness may differ both in leadership practices and in the way they are perceived by their subordinates.

The results of this study support the value of subgroup analysis in validating biographical inventories. None of the biographical inventories predicted success of officers assigned to Fleet billets. It may well be that different performance factors, personal qualities, and rater perceptions are involved in Shore billets as compared to Fleet billets. The fact that different kinds of items discriminated high from low criterion officers in the item analysis suggests such a possibility. This possibility is also supported by differences in the prediction of success in Shore and Fleet billets by the Cooperation Scale. The difference between r's of .30 and —.08 was significant beyond the .01 level of confidence. Research in non-Naval settings has also turned up evidence on this point; Vanasek (8) found that the personal characteristics required for success as a leader depend in part on the type of duty assignment.

A more parsimonious explanation for the differences in prediction in the two situations might be made in terms of the small samples and the known low reliability of biographical type items. Further, the fact that OCS Peer Ratings predicted equally well the performance evaluations of officers in Shore and Fleet billets suggests that there may be elements in common to these two duty assignments.

Such findings that a biographical inventory is valid in one situation and not in another has implications for further research. A job analysis may be needed to identify the situational requirements for effective leadership in the nonpredicted group. A job analysis also may help to develop a rationale for item selection and *a priori* scoring of new biographical inventories.

There is a need to continue to conduct research on the selection and assignment of leaders according to their preferred style of leadership. The style of leadership that an individual implements is not something he can turn on and off. It is likely to reflect various personality patterns and values. Furthermore, a leader is more likely to be effective following a style in which he feels most comfortable.

E. Summary

The best predictions of success as an officer on the job were obtained with self-descriptive biographical inventories measuring cooperativeness and a tendency to prefer a leadership style that "gets things done" rather than "human relations."

References

1. Fiedler, F. E. Engineer the job to fit the manager. *Harvard Bus. Rev.*, 1965, **43**, 115-122.
2. Katz, D., Maccoby, N., & Morse, N. Productivity, Supervision and Morale in an Office Situation: Part 1. Ann Arbor, Mich.: Univ. Michigan Press, 1950.
3. Maher, H. Construction of a forced-choice cooperation test and an investigation of matching indices of potential value to forced-choice questionnaires. Undated publication, Iowa State College, Ames, Iowa.
4. Moore, J. V. Factor analytic comparison of superior and subordinate ratings of the same NCO supervisors. *Tech. Rep. 53-24,* Human Resources Research Center, Randolph Air Force Base, San Antonio, Texas, 1953.
5. Owens, W. A. In E. R. Henry (Chairman), *Research Conference on the Use of Autobiographical Data as Psychological Predictors.* Conference report, The Creativity Research Institute of The Richardson Foundation, Greensboro, North Carolina, June, 1965, p. 78.
6. Pelz, D. C. Influence: A key to effective leadership in the first-line supervisor. *Personnel,* 1952, **29**, 209-217.

7. Torrance, E. P., & Ziller, R. C. Risk and life-experience: Development of a scale for measuring risk-taking tendencies. *Res. Bull. AFPTRC-TN-23,* Air Force Personnel and Training Center, Stead Air Force Base, Reno, Nevada, 1957.
8. Vanasek, F. J. An officer effectiveness report blank: Its exploitation and potential. Paper read at the Triservice Conference on Selection Research, Pensacola, Florida, May, 1960.
9. Williams, W. E. Life history antecedents of volunteers vs. non-volunteers for an A.F.R.O.T.C. program. Paper read at Midwestern Psychological Association Convention, Chicago, Illinois, 1961.

Department of Psychology
University of Windsor
Windsor, Ontario, Canada

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 115-118.

UNTIMED NORMS FOR THE SRA MECHANICAL KNOWLEDGE AND SHOP ARITHMETIC TESTS*

Student Counseling Center, University of Wisconsin; and Psychology and Student Counseling Service, Iowa State University

KENTNER V. FRITZ AND MARTIN F. FRITZ

A. INTRODUCTION

Under commercial conditions it is frequently desirable to administer tests promptly to individuals as they make applications for jobs. When time-limit tests are administered, there is not only a chance of error in timing but the clerical worker administering the tests, because of divided attention, cannot work efficiently at some other task. Moreover, a time-limitation when taken seriously by the applicant must often result in a certain amount of anxiety and unhappiness as indicated by the frequently heard remark, "I think I could have done better if I would have had a little more time!" And finally, from the standpoint of the test-user, it may be desirable to have every candidate show what he can do on a total body of content. Speed may not always be the primary consideration.

B. PROCEDURE AND RESULTS

We are presenting the test results of unselected, noncollege men, 80 percent high school graduates, who made application for various jobs requiring some degree of mechanical interest and knowledge such as truck driver, heavy equipment operator, equipment clerk, meter repairman, firefighter, power plant maintenance man, etc. All had been given the Otis Quick-Scoring Mental Ability Test, Gamma (OG-*IQ*), 30 minutes; and the Science Research Associates (SRA) Mechanical Aptitudes test, Form AH (1), Mechanical Knowledge (MK) and Shop Arithmetic (SA) but not the Space Relations, administered without time limit. Although the authors designed the SRA test battery primarily for prediction of training potential, this was not the objective of the test administration in this study. The purpose was to get an estimate of the present state of knowledge of each applicant with respect to commonly encountered tools and mechanical devices, and competence in simple applied arithmetic. The tests may be considered as having good content

---

* Submitted through John A. Antoinetti of the Editorial Board; received in the Editorial Office, Provincetown, Massachusetts, on March 3, 1969; and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

TABLE 1
AGE DISTRIBUTION OF MEN FOR THE MECHANICAL KNOWLEDGE AND SHOP ARITHMETIC TESTS

Age	MK $N = 271$	SA $N = 177$
18-20	26	18
21-25	89	60
26-30	49	35
31-35	27	24
36-40	17	13
41-45	21	10
46-50	21	8
51-55	12	4
56-60	9	5

validity for this purpose, on the basis of which selections may be made for employment.

The age distributions of men who took the MK and the SA tests are shown in Table 1. The OG-*IQ* distributions are shown in Table 2.

Although the distribution for MK, Table 3, shows a considerable skew toward higher scores, this test is of value in making a judgment as to whether or not an individual has acquired the usual fund of information possessed by others concerning general mechanical situations. In practice, it has been found that those making the lower scores on this test may profess an interest in mechanical situations, but have not actually done much about it. The SA test, Table 4, resulted in a more balanced distribution.

When the raw scores on the two tests were plotted against age and the

TABLE 2
OTIS GAMMA *IQ* DISTRIBUTION OF MEN FOR THE MECHANICAL KNOWLEDGE AND SHOP ARITHMETIC TESTS

IQ	MK $N = 271$	SA $N = 177$
131-135	8	4
126-130	17	10
121-125	19	10
116-120	29	16
111-115	19	13
106-110	31	28
101-105	33	21
96-100	37	30
91-95	29	20
86-90	15	9
81-85	14	8
76-80	11	5
71-75	9	2
66-70	0	1

OG-*IQ*s, it was found that the distributions deviated markedly from normality. For this reason it was deemed advisable not to calculate Pearson Product-Moment correlations but to make use of the Rank Differences method. The MK compared with age and OG-*IQ* gave Rho values of .09 and .31, respectively. The equivalent values for SA are .07 and .67. From this it may be inferred that the OG-*IQ* is related to the two tests, but to a much less extent in the case of MK. The moderately strong Rho value in the case of SA may indicate that, in spite of the SA being a pictorial test and supposedly so applied, there still is in it a rather strong element of what is often referred to as "mental ability."

TABLE 3
MECHANICAL KNOWLEDGE: SCORE, FREQUENCY, AND PERCENTILE DISTRIBUTION
($N = 271$)

Score	*f*	Percentile
45	4	99
44	7	98
43	15	96
42	20	90
41	35	83
40	25	70
39	28	61
38	24	51
37	14	42
36	19	37
35	17	30
34	11	23
33	9	19
32	4	16
31	4	14
30	4	13
29	4	11
28	5	10
27	6	8
26	2	6
25	2	5
24	3	4
23	2	3
22	3	3
19	1	2
15	1	1
10	1	1
0	1	1

An increase in scores obtained by time-limit testing is reported by SRA for both MK and SA from the ninth grade through the 12th. In this study, practically no relation with age was found even though a cumulative effect of experience would be expected. Although the explanation might be given that high-scoring older men, having long before found stable employment, were not well represented in the samples, the tests still constitute a face-valid method of handling the practical problem of employee selection, hopefully to be verified by later research.

TABLE 4
SHOP ARITHMETIC SCORE, FREQUENCY, AND PERCENTILE DISTRIBUTION
($N = 177$)

Score	f	Percentile
24	1	99
23	5	99
22	5	97
21	5	94
20	10	91
19	8	85
18	9	81
17	12	76
16	11	69
15	14	63
14	19	55
13	12	44
12	10	37
11	11	32
10	16	25
9	6	14
8	6	13
7	9	10
6	2	5
5	2	3
4	4	2

REFERENCE

1. SRA MECHANICAL APTITUDES. Grades 9-12 and adults. Chicago, Ill.: Science Research Associates, 1947, 1950.

Student Counseling Center
The University of Wisconsin
Madison, Wisconsin 53706

Department of Psychology
Student Counseling Service
Iowa State University
Ames, Iowa 50010

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 119-124.

THE ROLE OF VERBAL FACTOR IN A FORCED-CHOICE TEST OF ACHIEVEMENT VALUE*[1]

Department of Psychology, York University, Canada

BISHWA NATH MUKHERJEE

A. INTRODUCTION

Although it has been observed by many that the verbal ability of the *S* has a profound impact upon how he responds in a personality questionnaire, only a few empirical studies have been reported in this connection. Seigman (14) found, for example, a negative correlation between scores on the vocabulary test of the Shipley-Hartford Retreat Scale and the California F Scale. The correlation was found to be —.38, significant at .05 level. Messick and Kogan (6) in a factor-analytic study also isolated in their male sample a verbal comprehension factor marked at one end by a vocabulary test with a loading of .71 and the other end by a modified F Scale with a loading of —.45.

With Taylor Manifest Anxiety Scale (MAS), the results have been somewhat contradictory. Calvin, Koons, Bingham, and Fink (1) reported a significant negative (—.29) correlation between MAS and the Vocabulary subtest of Wechsler Bellevue Intelligence Test. Kerrick (4) also found that scores on Word Knowledge Test of AAF battery correlated .40 with MAS in a sample of 128 Air Force trainees. Matrazzo, Ulett, Guze, and Saslow (5), however, had shown that composite scores derived from the Comprehension, Vocabulary, and Similarity subtests of the Wechsler Bellevue Scale, given in group administration to 101 college sophomores, correlated —.07 only with the MAS scores. Mukherjee (8) reported a correlation of .004 and .065 with MAS scores on the one hand and the Word Ending and Word Beginning Tests, respectively. It appears then that scores on the MAS may be relatively more independent of both verbal comprehension and verbal fluency factors than the California F Scale.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on March 6, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The research reported herein was supported through the Cooperative Research Program of the Office of Education, United States Department of Health, Education, and Welfare during the author's period of stay at Indiana University.

In the area of questionnaire measurement of achievement-related needs (v Ach), not a single study has been reported so far which deals with the role of verbal factors. The present study was taken up for assessing the importance of two verbal factors in determining the v Ach scores of American college students. Specifically, the object of the study was to determine the extent poor v Ach scores are due to vocabulary limitations on the part of respondents.

B. METHOD

1. *Subjects*

The *S*s were 87 freshmen enrolled in an introductory course offered during fall 1963 semester at the Southeastern Campus of Indiana University, Jeffersonville, Indiana. A detailed description of the sample appears elsewhere (9).

2. *Tests*

Two measures of verbal ability—namely, vocabulary and completion—were taken from the original 57 tests of Primary Mental Abilities (PMA) used by Thurstone (15) and published by him as a supplement (16). In addition to these tests, Thurstone and Thurstone's (17) Suffixes and Prefixes as adapted recently by Educational Testing Service, Princeton, New Jersey, under the titles Word Endings and Word Beginnings, respectively (3) were used as predictors. These two tests are supposed to measure the word fluency factor (3), since both require the *S*s to produce rapidly words that fulfill certain requirements.

For measuring achievement values (v Ach), the Sentence Completion Test (SCT) developed by Mukherjee (7, 8) was used. The SCT consists of 50 forced-choice triads in which items having comparable social desirability values based on the ratings of American college students samples are arranged in groups of three so that each member of the triad represents a different psychological need. Each triad, however, has an item the endorsement of which reflects achievement orientation. The total number of such achievement related items endorsed by the *S* operationally defines his v Ach. A high score on the SCT is interpreted as a keen desire to compete successfully with a standard of excellence, an expressed interest in undertaking difficult and challenging tasks, and a strong sense of optimism. Details concerning the construction, administration, scoring procedure, and various psychometric properties of SCT appear elsewhere (7, 8). The Kuder-Richardson reliability estimate for SCT has been found to be .716 for a sample of 248 students enrolled in Indiana

University. The test retest reliability of the SCT after an interval of nearly two months has been found in various studies to range from .71 to .83. Studies demonstrating the concurrent and construct validities of SCT as a measure of achievement value have been reported elsewhere (8, 10, 11, 12, 13).

3. *Procedure*

Mimeographed copies of SCT were distributed to all the *S*s and they were urged to fill out and return the test booklet and the completed answer sheet within a week. All the *S*s filled out the questionnaire in their leisure time in accordance with the instructions contained on the first page of SCT booklet, and turned in the material within three days. In addition to SCT, each member took a large number of psychological tests including 20 tests from Thurstone's PMA, the Word Beginning, and Word Ending Tests and several other ability tests. These tests were administered during the regular class hours between November 13, 1963 and December 17, 1963. These tests including the four verbal tests used for the present study were administered in groups of nearly 40 *S*s at a time, as a part of a larger project on learning and structure of abilities. In addition to the investigator himself, one research assistant was responsible for the administration and proctoring of all tests. Since the whole group agreed to take the tests on the condition that each participant will be counseled later on, there was evidence of an optimal motivation on the part of *S*s. Not more than four tests were administered at a time, and in each case there was at least a five-minute rest between two tests. The testing sessions were held in the classroom which provided optimal conditions for the taking of psychological tests.

C. RESULTS

The product moment intercorrelations among the verbal tests, as well as SCT, are reported in Table 1 along with the mean and standard deviation of each variable. Multiple correlation between all the four verbal tests on the one hand and the SCT on the other hand was also computed. The multiple correlation turned out to be .26; and, thus, the hypothesis of zero correlation between the verbal predictors and SCT could not be rejected.

D. DISCUSSION

The results reported in this paper indicate clearly that v Ach scores on the SCT are neither determined by the verbal fluency nor by the vocabulary limitations on the part of the respondents. It appears that the level of vocabulary used in making the items of SCT was relatively easier for the sample of

American college freshmen studied. However, it should not be interpreted that the items of SCT would have necessarily the same meaning for all *S*s who took the test. This may or may not be true. The assumption of an identical meaning of a particular item across the population of *S*s is not particularly involved in the construction of SCT.

Since "the word fluency factor W is involved whenever the subject is asked to think of isolated words at a rapid rate" (17, p. 110), as is called for in a test of rhyming, or in reproducing words with a given letter, prefix, or suffix, it is no wonder that SCT scores showed almost a zero correlation with scores on Word Beginning, as well as Word Ending Tests.

TABLE 1
PRODUCT MOMENT INTERCORRELATIONS[a] AMONG SELECTED VERBAL TESTS OF PMA AND SENTENCE COMPLETION TEST ALONG WITH THE MEAN AND STANDARD DEVIATION OF EACH VARIABLE
($N = 87$)

Variable	SCT	Completion	Vocabulary	Word Ending	Word Beginning
Sent. Completion Test	1.000	.200	.172	.007	—.066
Completion	.200	1.000	.545	.269	.287
Vocabulary	.172	.545	1.000	.339	.317
Word Endings	.007	.269	.339	1.000	.548
Word Beginnings	—.066	.287	.317	.548	1.000
Mean	15.46	21.82	40.63	28.43	23.01
Standard deviation	4.88	5.88	11.51	7.62	6.05

[a] In order to be significant at .05 level, the required value is .21.

It is of supplementary interest to note the positive but quite low correlations between scores on Vocabulary Test and the two measures of word fluency used in the present study. The results indicate that, although the Word Beginning and Word Ending Tests measure some kind of familiarity with a wide range of words, it is not the same kind of "knowledge" of words that is elicited by measures of verbal comprehension factor called V, such as the PMA Vocabulary Test. The distinction between these two verbal factors has been noted from time to time in various factor-analytic studies (2, p. 198).

The results reported here appear to indicate that the v Ach scores of SCT have little verbal saturation in the population of American college freshmen. The notion that *S*s having high achievement values will tend to have a more enriched vocabulary than *S*s with low v Ach scores is not supported by the data of the present study. The study also shows that achievement orientation of American college freshmen cannot be predicted if we know about their strength and weakness in the verbal domain. To what extent such a generalization will be true for other types of populations merits further study.

E. SUMMARY

In order to investigate if scores on Sentence Completion Test (SCT), a forced-choice scale of achievement value (v Ach), are dependent upon verbal fluency and vocabulary, 87 freshmen studying at Indiana University, Jeffersonville, were given the Word Beginning, Word Ending, and Vocabulary Tests after they took SCT. The multiple correlation between scores on SCT and verbal tests turned out to be a nonsignificant .26. It appears that *S*s' verbal fluency and vocabulary play a negligible role in determining their scores in SCT.

REFERENCES

1. CALVIN, A. D., KOONS, P. B., JR., BINGHAM, J. L., & FINK, H. H. A further investigation of the relationship between manifest anxiety and intelligence. *J. Consult. Psychol.*, 1955, **19**, 280-282.
2. ERLEBACHER, A., & HARRIS, C. W. Parameters of word fluency tasks. *J. Educ. Psychol.*, 1962, **53**, 198-202.
3. FRENCH, J. W., *Ed.* Kit of Selected Tests for Reference Aptitude and Achievement Factors. Princeton, N.J.: Educ. Testing Service, 1954.
4. KERRICK, J. S. Some correlates of the Taylor Manifest Anxiety Scale. *J. Abn. & Soc. Psychol.*, 1955, **50**, 75-77.
5. MATRAZZO, J. D., ULETT, G. A., GUZE, S. B., & SASLOW, G. The relationship between anxiety level and several measures of intelligence. *J. Consult. Psychol.*, 1954, **18**, 201-205.
6. MESSICK, S., & KOGAN, N. Categorizing styles and cognitive structure. Research Bulletin, Educational Testing Service, Princeton, New Jersey, 1967.
7. MUKHERJEE, B. N. Development of a forced-choice test of achievement motivation. Final Report, Project No. 113, U.S. Office of Education, and Indiana University, Bloomington, Indiana, August, 1964.
8. ———. A forced-choice test of achievement motivation. *J. Indian Acad. Appl. Psychol.*, 1965, **2**, 85-92.
9. ———. The prediction of grades in introductory psychology from tests of Primary Mental Abilities. *Educ. & Psychol. Meas.*, 1965, **35**, 557-564.
10. ———. Achievement motivation and goal setting behaviour in the classroom. *Brit. J. Educ. Psychol.*, 1965, **35**, 286-293.
11. ———. Self perception and achievement motivation. *J. Psychol. Res.*, 1965, **9**, 109-114.
12. ———. Achievement values, social desirability and endorsement of trait names on the Berdie Check List. *J. Appl. Psychol.*, 1968, **52**, 127-132.
13. ———. Achievement values and scientific productivity. *J. Appl. Psychol.*, 1968, **52**, 145-147.
14. SEIGMAN, A. W. The effect of cultural factors on the relationship between personality, intelligence and ethnocentric attitudes. *J. Consult. Psychol.*, 1958, **22**, 375-377.

15. Thurstone, L. L. Primary mental abilities. *Psychomet. Monog.*, 1938, No. 1.
16. ———. Primary Mental Abilities (Supplement). Chicago, Ill.: Univ. Chicago Press, 1938.
17. Thurstone, L. L., & Thurstone, T. G. Factorial studies of intelligence. *Psychomet. Monog.*, 1941, No. 2.

Department of Psychology
York University
Toronto 12, Ontario, Canada

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 125-129.

AN INVESTIGATION OF SUCCESS PREDICTORS IN GRADUATE SCHOOL IN PSYCHOLOGY*

Institute of Behavioral Research, Texas Christian University

JAMES R. RAWLS, DONNA J. RAWLS, AND C. WADE HARRISON

A. INTRODUCTION

In selecting students for graduate work in psychology, most departments usually use some combination of GRE or MAT scores, undergraduate grade point average, and letters of recommendation. Yet, from a practical standpoint, these measures are predictively weak, accounting for only a small proportion of the variance in the criterion measures (1).

Obviously, one of the reasons for the failure of these predictive measures to demonstrate later discriminative efficiency is the very fact that they are used in the selection process. In effect, by eliminating students who were less successful on the predictive measures, much of the relevance to later criterion measures is removed. Truncating the distribution in this manner and then making predictions on the resultant homogeneous sample could hardly be expected to produce impressive predictive validity coefficients.

Still another problem is that of adequate criterion measures. Studies to date have utilized comprehensive examinations, faculty ratings, student ratings, and graduate grade point average; yet, ratings, including those of performance on comprehensive exams, have demonstrated far less than optimal reliability. Furthermore, even without considering the subjectivity in assigning grades, the limited range (usually only A's, B's, and C's) restricts their capacity for prediction. Moreover, obtaining a Ph.D. in psychology is dependent upon considerably more than passing an exam or making good grades: e.g., performance in oral examinations, successful completion of a thesis, and successful completion of a dissertation.

Regardless of the problems, there is a greater need for predictors of success in graduate training today than ever before in history. There is an increased demand for advanced degrees, and there are ever-increasing numbers of applicants to fill available slots. Therefore, concentrated attention will need to be focused upon resolving these problems and finding ways and means

---

* Received in the Editorial Office, Provincetown, Massachusetts, on March 7, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

whereby we can make better and more sophisticated predictions of success in graduate training than we have done in the past.

The purpose of the present study was to investigate the utility of a number of predictor variables, both new and old, in discriminating between three groups of students in psychology at Louisiana State University: viz., those obtaining Ph.D.'s, those receiving terminal master's degrees, and those who were unable to obtain either degree at L.S.U.

B. METHOD

1. *Subjects*

The subject population consisted of 75 students who were enrolled in the graduate program in psychology at Louisiana State University between September, 1961, and August, 1965. Of these, 25 subjects obtained the Ph.D. during this time, 25 obtained a terminal master's degree, and 25 were unable to obtain either degree. This latter group was not allowed to continue in the program after their first year because of academic failure. There were 65 males and 10 females, ranging in age from 20 to 50.

2. *Procedure*

Data collected included MAT and GRE scores, letters of recommendation, a few biographical items, and an essay answering the question, "Why I want to go to graduate school in psychology." In addition, transcript information was recorded in order to show the number of hours taken within each subject area and the number of hours within each area of A's, B's, C's, etc. Thus, in addition to overall grade point average, grade point averages were calculated for each subject area.

C. RESULTS

The data were analyzed by means of appropriate t's and χ^2's. As shown in Table 1, results of the present study indicated that those *S*s who obtained the Ph.D. had had more undergraduate hours in psychology, biological sciences, and math, and fewer hours in the humanities than either terminal master's *S*s or *S*s not obtaining a graduate degree. Ph.D. *S*s had had more hours in foreign languages than M.A. *S*s, but not more than the no degree *S*s. Ph.D. *S*s and terminal master's *S*s had taken more undergraduate hours in the natural and physical sciences and combined sciences than *S*s not obtaining graduate degrees. The number of hours in the natural and physical sciences and combined sciences were the only two significant differences that were found in the entire study between M.A. and no degree *S*s.

*S*s who received the Ph.D. had higher grade point averages in humanities, math-science courses, and in the combined sciences. Ph.D. *S*s had higher grade point averages in the natural and physical sciences than no degree *S*s, but not higher than master's degree *S*s.

TABLE 1
SIGNIFICANCE LEVELS OF DISCRIMINATING PREDICTOR VARIABLES

Variable	Ph.D.-M.A.	Ph.D.-No degree	M.A.-No degree
Hours in			
Humanities*	$<.20$	$<.20$	
Foreign languages	$<.10$		
Natural and physical sciences		$<.20$	$<.20$
Mathematics	$<.20$	$<.20$	
Biological sciences	$<.20$	$<.10$	
Psychology	$<.05$	$<.05$	
Combined sciences		$<.20$	$<.20$
Grade point average in			
Humanities	$<.20$	$<.10$	
Natural and physical sciences		$<.05$	
Math-science	$<.20$	$<.05$	
Combined sciences	$<.20$	$<.20$	
Biographical information			
Age at admission	$<.01$	$<.01$	
Married at admission	$<.001$	$<.001$	
Entering with M.A.	$<.01$	$<.01$	
Large undergraduate university	$<.10$	$<.10$	

* Ph.D.'s and M.A.'s had higher means or frequencies in each case except the humanities.

With regard to the biographical items, *S*s who obtained a Ph.D. as compared with the other *S*s were older when entering the program, were more often married upon admission, and more likely entered with a master's degree. Ph.D. *S*s also came from larger undergraduate institutions than did the *S*s in the other two groups.

In an attempt to make some predictive use of letters of recommendation, use was made of a variety of procedures ranging from counting the number of positive adjectives and adverbs to having judges rate the letters on various criteria. None of these procedures was effective in differentiating the three groups. Similarly, a number of techniques were tried in order to make some predictive sense out of the essay answering the question, "Why I want to go to graduate school in psychology," but to no avail.

Despite the large number of variables investigated, no other variables differentiated the three groups. Variables which failed to reach significance are listed in Table 2.

TABLE 2

VARIABLES FAILING TO DISCRIMINATE

- Test data
  - MAT
  - GRE-Verbal
  - GRE-Quantitative
  - GRE-Total
  - GRE-Psychology
- Hours in
  - Art
  - Education
  - English
  - Music
  - Philosophy
  - Physical education
  - Sociology
  - Social science
  - Theology
  - Combined sciences and math
- Grade point average in
  - Art
  - Biological science
  - Education
  - English
  - Languages
  - Math
  - Music
  - Philosophy
  - Physical education
  - Psychology
  - Sociology
  - Social science
  - Theology
  - Combined psychology and social science
  - Overall
- Location of BA/BS conferring school
  - N.-S.
  - E.-W.
  - S.-not S.
  - NE-SE-NW-SW
- Degree upon entrance
  - B.S.-B.A.
  - M.S.-M.A.
- Undergraduate major
  - Psychology-nonpsychology
  - Social science-nonsocial science
- Sex

D. DISCUSSION

Despite a rather large number of variables and an unusually low acceptable level of confidence ($p < .20$), few variables showed promise in their capacity to discriminate successful, less successful, and unsuccessful graduate students in psychology. As indicated, both the MAT and GRE failed significantly to differentiate the three groups. Likewise, letters of recommendation and the essay on "Why I want to go to graduate school in psychology" did not discriminate between the groups.

The partitioning of transcript information did show some glimmer of usefulness in that Ph.D.'s, as compared with the other *S*s, had had more undergraduate hours in psychology, foreign languages, biological sciences, math, natural and physical sciences, and all sciences combined. They also had higher grade point averages than the other *S*s in the humanities, natural and physical sciences, math-science courses, and combined sciences.

Biographical information also showed some encouraging predictive value in that Ph.D. *S*s were older than the other *S*s when entering the program, were more often married upon admission, and more frequently entered with a master's degree. They also more often obtained their undergraduate degrees from a larger university than did the other *S*s.

As indicated by their overall undergraduate point average, MAT and GRE scores, graduate students accepted into the L.S.U. psychology program were very homogeneous with regard to their intellectual ability. It would thus appear that success in graduate training was dependent upon some factor other than sheer brain power: e.g., interests, motivation, or personality. But whatever the factor may be, it seems fitting to look for predictors outside the realm of intellectual ability so long as students are carefully screened before entering the program. Results of this study indicate that past experience in course work areas, and biographical information (age, marital status, etc.) appear to have some predictive potential. In light of recent successes in personnel selection in industry with the use of the Biographical Information Blank, the Personal History Survey, and similar devices, biographical data may prove to be a very fruitful area of investigation for predicting success in graduate training.

REFERENCE

1. ROBERTSON, M., & HALL, E. Predicting success in graduate study. *J. Gen. Psychol.,* 1964, **71,** 359-365.

Institute of Behavioral Research
Texas Christian University
Fort Worth, Texas 76129

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 131-134.

THE EFFECT OF SAMPLE ITEMS ON TEST PERFORMANCE*

Institute for Child Study, University of Maryland; and
Department of Education, Purdue University

Donald W. Felker, Gerald Larson, Frank Milhollan,
and John R. Pyper

A. Introduction

It has been found that questions inserted in filmed or printed instructional materials facilitate the learning of the information covered by the inserted questions and also the learning of material untested during the study period (1, 2, 4, 6, 7).

It should be noted that the generally facilitative effect on learning associated with inserted questions only occurs under certain conditions. If the questions are seen just before the relevant material is read *vs.* just after, or if the answers to the questions are given, only the ability to answer the inserted questions is improved (1, 6, 7). If the material is quite interesting, inserted questions apparently have little or no effect (4), as well as if the important information in the text is cued by underlining, printing in red, etc. (2, 3).

Additional research findings suggest that there are conditions under which the facilitative effect of inserted questions is selective. *S*s exposed to inserted questions on specific types of information, such as common and technical phrases or measures and names, tended to learn more of the type of information involved in the inserted questions (7).

In spite of the conclusions of the above mentioned researchers concerning classroom and textbook procedures, Pyper (5) has argued that the findings are as yet of dubious value for practical application because (*a*) the learning tasks of the research activities have been short instructional experiences which are atypical of the learning atmosphere and environment of the classroom, (*b*) the task is not course imbedded and hence raises the question of the Hawthorne effect, and (*c*) the research instructional procedure of not allowing the *S*s to backtrack and reread material is unrepresentative of practically all classroom instruction.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on March 10, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

B. Problem

The present study was designed to investigate the effect on student attitude and learning of handing out sample test items when the students are preparing for a test. This approach imbeds the inserted question procedure in an ongoing instructional program. The need to prepare for a classroom test should serve as motivation for mastery of material.

The following hypotheses were tested:

1. *S*s given sample items for study prior to a test will correctly answer more test questions involving the material covered by the sample items than the material not covered by the sample items.
2. *S*s given sample test items for study will perceive the procedure as being helpful in preparing for the test.

C. Method

1. *Subjects*

Students enrolled in a master's level course in "Learning Theory" at the University of Maryland were selected for the study. Two sections of 28 students each served as subjects.

2. *Procedures*

A midterm test of 112 items covering four chapters in the required textbook was prepared. The test was then given to Section I of the course. Item analysis data from the responses of the students in Section I was used and two tests were compiled. One test consisted of 25 items from chapters 1 and 3. A second test consisted of 25 items from chapters 2 and 4. With the use of the item analysis data, items on the two tests were matched for difficulty level and item discrimination.

Items not included in the two 25-item matched-tests, and only covering material in chapters 1 and 3, were handed out to Section II of the course. These items were presented as sample items to prepare the students for the test and to aid them in studying. One week later the two matched-tests of 25 items each were combined into a 50-item test and given to the students in Section II as a midterm examination. After the test each student answered the following question: "Did you feel that the test questions distributed to you were helpful? Very much. Quite a bit. Not much. Not at all."

The differences between the scores on the two matched tests were compared in testing the first hypothesis that *S*s given sample items for study prior to a test will correctly answer more test questions involving the material covered by the sample items than the material not covered by the sample items. The analysis is presented in Table 1.

TABLE 1
ANALYSIS OF VARIANCE OF SCORES ON A TEST OVER MATERIAL WITH SAMPLE TEST ITEMS AND A TEST OVER MATERIAL WITH NO SAMPLE ITEMS

Source	*SS*	*df*	*MS*	*F*
Between subjects	992.9	27	36.7	12.655
Between tests	18.3	1	18.3	6.310*
Error	79.7	27	2.9	
Total	1090.9	55		

* $p < .05$.

D. RESULTS

The analysis indicates that the students scored significantly higher on the test covering chapters from which sample items were taken ($\overline{X} = 15.42$) than on the test covering chapters from which no items were taken ($\overline{X} = 14.28$). The *F*-ratio for tests was 6.31 with 1 and 27 degrees of freedom, $p < .05$.

The performance on each of the two sets of test items of the *S*s who received sample test items and those who did not receive sample items was compared. Group one, which had no sample items, had the following proportions of correct responses on the tests: (*a*) test over chapters one and three, 59.68; (*b*) test over chapters two and four, 59.84; and (*c*) total for combined tests, 59.76. Group two, which had sample items from material in chapters one and three, had the following proportions of correct responses on the tests: (*a*) test over chapters one and three, 61.71; (*b*) test over chapters two and four, 57.14; and (*c*) total for combined tests, 59.42.

It is apparent that there is no difference in the total proportion of items answered correctly for either condition. Thus the rise in the proportion of correct answers to items which covered the material represented in the sample items was accompanied by a decrease in the proportion of correct answers to items which covered information not represented in the sample items.

In order to test the second hypothesis that *S*s given sample test items for study will perceive the procedure as being helpful in preparing for the test, the distribution of responses to the posttest question concerning the helpfulness of the sample items was analyzed with the use of chi square. The number of students choosing each response was Very Much, 12; Quite a Bit, 10; Not Much, 2; and Not at all, zero. Four students did not mark their answer sheets.

A random distribution for expected frequencies was assumed and the chi square equals 17.333, which has a $p < .01$.

E. SUMMARY AND DISCUSSION

The results can be summarized as follows: significantly more students said that sample items are helpful than students who said they are not helpful. Giving students sample items was associated with higher performance on test items covering material from which sample items were taken. Finally, students who had sample items did not attain an overall test achievement which was any higher than students who had no sample items.

The generalized yet selective facilitative effect of the sample test items is consistent with the finding of Rothkopf and Bisbicos (7). Such findings suggest that sample items result in increased posttest ability on material from which the items were taken but decreased posttest ability on material from which no items were taken. The findings of the present study would indicate that sample items before a test can affect learning, but that concern should be given to how the material in the sample items is distributed among the total material to be studied.

REFERENCES

1. FRASE, L. T. Learning from prose material: Length of passage, knowledge of results, and position of questions. *J. Educ. Psychol.*, 1967, **58**, 3-10.
2. HERSHBERGER, W. A. Self-evaluational responding and typographical cueing: Techniques for programming self-instructional reading materials. *J. Educ. Psychol.*, 1964, **55**, 288-296.
3. HERSHBERGER, W. A., & TERRY, D. F. Typographical cueing in conventional and programmed texts. *J. Appl. Psychol.*, 1965, **49**, 55-60.
4. MACCOBY, N., MICHAEL, D. N., & LEVINE, S. Further studies of student participation procedures in film instruction: Review and preview, covert practice, and motivational interactions. In Lumsdaine, A. A. (Ed.), *Student Response in Programmed Instruction*. Washington, D.C.: Nat. Acad. Sci., 1961.
5. PYPER, J. R. A review of the effects on learning of questions inserted in textual information. Mimeographed paper presented at the annual meeting of the National Society for Programmed Instruction, San Antonio, Texas, 1968.
6. ROTHKOPF, E. Z. Learning from Written Instructive Material: I. An Exploration of the Control of Inspection Behavior by Test-Like Events. Murray Hill, N.J.: Bell Telephone, 1965.
7. ROTHKOPF, E. Z., & BISBICOS, E. E. Selective facilitative effects of interspersed questions on learning from written materials. *J. Educ. Psychol.*, 1967, **58**, 56-61.

Department of Education
Purdue University
Lafayette, Indiana 47907

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 135-139.

COMPARISON BETWEEN AMERICAN PERMANENT RESIDENTS OF ISRAEL: PART II, ISRAELI BACKGROUND*[1]

B'nai B'rith Hillel Foundation at Purdue University

GERALD ENGEL

A. INTRODUCTION

Native American Jews who became permanent residents of Israel during the five-year period 1962 through 1966 provided information in 1967 about their American background and adjustment to Israel. Their responses relating to life in America were presented in Part I, as a comparison between the religiously oriented and the nonreligious.[2] Included in Part I was an analysis of the minority's reaction to a majority culture, identification with American Jewish life, and interest in Israel (2).

Both groups were equally disturbed by materialism in America. To a lesser degree the religious and nonreligious were equally troubled by anti-Semitism. The nonreligious were more disturbed by conformity and political witch-hunting; while the more Jewishly group-identified religious subjects were more concerned about the possibility of assimilating. The religiously oriented were also more interested in Israel than were the nonreligious.

Part II maintains the same religious, nonreligious comparisons as in Part I. Part II deals with subjects' attractions to Israel, feeling upon arrival and afterwards, and their comparison between working conditions in Israel and America. The American Jew who immigrated to Israel chose to reduce conflicts between belief and action by aliya (immigration). By contrast the

---

* Received in the Editorial Office, Provincetown, Massachusetts, on March 10, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The author gratefully acknowledges the aid of Dr. Harold Kahn, National Institute of Health, in preparing the questionnaire; of Dr. Ben Winer, Purdue University, for statistical analysis; Glenda McCracken and Arnold Woodruff, Purdue Computer Center, for programming. Marilyn Kopelowitz Engel served as study coordinator.

The author makes special acknowledgement to B'nai B'rith Hillel Foundations for granting him a decennial leave to gather the data, November 1966 through August 1967, and to Dr. Frederick L. Hovde, President of Purdue University, for allocating President's Funds for computer work.

[2] In Israel 255 subjects identified with Jewish religious tradition: 87 traditionalists (masorti); 30 keepers of commandments (shomer mitzvot); 138 observant (dati). One hundred eighty-eight were designated nonreligious in Israel: 27 against observance (antidati); 80 not observant (lo dati); 55 secularists (chiloni); 24 who wrote "don't know"; two Hebrew Christians.

American "alienated by his own commitment" (1) who remains in his native land generally modifies his own attitude to fit into the system. The individual who goes to live in Israel is responding to an inner signal to avoid increasing dissonance (3). Subjects in conflict about aspects of American life and about Israel's attraction are discussed in terms of an expanded theory of cognitive dissonance.

B. SUBJECTS

From an official list of permanent residents of Israel the names of 780 Americans born in the United States or Canada were selected, comprising all native Americans who became permanent residents of Israel during the years 1962 through 1966, who were 20 years of age or older by the end of 1966. Seventy-seven subjects were eliminated, including six who said they left America as infants, six who were deceased, and 65 who were not in Israel from March through August 1967 when this questionnaire was administered. Of 703 potential subjects 443 cooperated, representing a 63 percent response.

C. PROCEDURE

All subjects were mailed a four-page, self-administering questionnaire. Respondents included 364 who voluntarily answered one of four mass mailings and 79 other subjects who were considered forced, responding to personal letter or interview. When few statistical differences were observed between these groups, the subjects were combined. The comparison presented between religiously oriented and nonreligious is based upon subjects' religious alignment in Israel, and does not necessarily reflect their religious outlook when in America (2).[3]

D. RESULTS

Several hypotheses were postulated upon the assumption that the religiously oriented, more Jewishly identified in America, would be more attracted to, and experience greater satisfaction living in Israel than the nonreligious. One assumption was confirmed: the religiously oriented were more attracted to a Jewish Israel than were the nonreligious (Table 1).

Other conjectures were rejected. There was no group difference about career opportunities attracting settlers (Table 1). Both groups of olim (immigrants) before settling shared the same degree of optimism about living within Israeli income; 66.5 percent expected it to be easy.

---

[3] See footnote 2.

TABLE 1
ISRAEL'S ATTRACTION

Question	Percentages	
	Nonreligious (*N* 188)	Religiously oriented (*N* 255)
Homeland	(*N* 172)	(*N* 235)**
none	18.0	8.1
some	32.6	17.0
much	49.4	74.9
Land of Bible	(*N* 166)	(*N* 216)**
none	40.4	12.5
some	44.0	29.6
much	15.7	57.9
Jewish life	(*N* 167)	(*N* 230)**
none	31.1	8.3
some	41.3	17.8
much	27.5	73.9
Live among Jews	(*N* 170)	(*N* 230)**
none	24.7	8.7
some	32.4	21.7
much	42.9	69.6
Career opportunities	(*N* 164)	(*N* 205)
none	59.1	61.0
some	28.7	29.8
much	12.2	9.3
Jewish education	(*N* 167)	(*N* 317)**
none	41.9	18.4
some	23.4	23.5
much	34.7	58.1
Kibbutz life	(*N* 169)*	(*N* 203)
none	52.7	65.5
some	24.3	16.7
much	23.1	17.7

Note: All tables derived from BMD02s—Contingency Table Analysis—Version of March 11, 1964, Health Sciences Computing facility, University of California at Los Angeles.

* Chi square significant differences = .05.

** Chi square significant differences = .001.

There was no group difference in age of arrival to settle. Thirty-five percent came before age 25; and 34.3 percent between age 25 and 35.

There was a similar feeling of ease among both groups of settlers, encompassing 74.5 percent upon arrival, 84.3 percent a year later, and 95 percent at time of responding. Yet, there was and is a difference in ability to communicate in Hebrew, the language of the land. Upon arrival 54.5 percent of the nonreligious and 30.7 percent of the religious had no ability in conversational Hebrew. When responding 80.5 percent of the nonreligious and 93.5

percent of the religious report much ability in conversational Hebrew (chi square significance at .001 level).

The nonreligious working in Israel feel there is more opportunity in their fields when comparing Israel to America than do the religious. Nevertheless, there is similarity of feeling between groups about income, satisfaction, prestige, and unemployment in their field, as well as standard of living (Table 2).

Real occupational differences exist with the chi square significant difference at .01. More of the nonreligious are professionals (27.7 percent of the nonreligious compared to 22.0 percent of the religious) and farmers (9.0 percent of the nonreligious, 3.1 percent of the religious).

Among these immigrants 89.5 percent agree that their housing is adequate. More than half (57.3 percent) assert living within their income is easy.

The kibbutz was more attractive to the nonreligious than to the religious (Table 1). More of the nonreligious live in rural areas (27.7 percent of

TABLE 2
ISRAEL COMPARED TO AMERICA

	Percentages	
Question	Nonreligious (*N* 188)	Religiously oriented (*N* 255)
Opportunities in field	(*N* 143)*	(*N* 184)
less	48.3	44.0
same	23.1	37.0
more	28.7	19.0
Income for work	(*N* 139)	(*N* 175)
less	80.6	80.6
same	12.2	9.7
more	7.2	9.7
Satisfaction in work	(*N* 153)	(*N* 193)
less	19.0	19.7
same	35.3	32.6
more	45.8	47.7
Prestige in work	(*N* 148)	(*N* 179)
less	25.7	29.6
same	35.1	31.8
more	39.2	38.5
Standard of living	(*N* 172)	(*N* 215)
less	76.7	68.8
same	19.2	22.8
more	4.1	8.4
Unemployment in my field	(*N* 91)	(*N* 114)
less	28.6	21.1
same	50.5	50.9
more	20.9	28.1

* Chi square significant differences = .02.

the nonreligious, 16.5 percent of the religious at chi square significant difference .005). Of all American olim (immigrants) 21 percent live in rural areas compared to 11.6 percent for Israel (5) and between three and four percent of Jews in the U.S.A. (6).

E. Discussion

Religiously oriented American Jews were more attracted to a Jewish Israel than were the nonreligious. Nonreligious subjects found kibbutz life more appealing than did the religious. Both groups of immigrants, responding to many needs, settled in Israel at the same ages (4). Both groups were equally comfortable after arrival, a fact which was evident in their assertions about feeling at ease, adequacy of housing, ease of living within income, and job satisfactions.

The religiously oriented were more attracted to a Jewish Israel than the nonreligious, though the Jewish attractions of the land were paramount to both groups. Despite dissimilarities, the overall view of America and Israel created the same total dissonance within both groups. This was reflected in similarities of age when settling, and similar feelings of ease upon arrival, though differences existed between groups in ability to speak Hebrew. In 1967, both groups coincided in their views about housing, living within Israeli income, and work satisfactions, notwithstanding real differences in occupation, and expressed disagreement about occupational prospects.

References

1. Bloy, M. B., Jr. Alienated youth, their counter culture, and the chaplain. *Church Rev.*, Nov. 1968, No. 1, 8-14.
2. Engel, G. Comparison between American permanent residents of Israel, Part I: American background. *J. of Psychol.*, 1969, **71**, 133-142.
3. Kelman, H. C. Discrepant action and attitude change: A functional analysis. Dittoed colloquium presented at University of Oregon, Eugene, April 28, 1967; and at University of California, Santa Barbara, May 16, 1967.
4. Lewin, K. The background of conflict in marriage. In G. W. Lewin (Ed.), *Resolving Social Conflicts*. New York: Harper, 1948. Pp. 84-102.
5. Statistical Abstracts of Israel 1967 (Vol. 18). Jerusalem, Israel: Government Press, 1967. P. 27, Table B/7.
6. U.S. Dept. of Commerce, Bureau of the Census. Tabulations of data on the social and economic characteristics of major religious groups, March 1957. Mimeographed unpublished 10 pages, 20 tables, January, 1968.

B'nai B'rith Hillel Foundation at Purdue University
State and Waldron
West Lafayette, Indiana 47906

Journal of Genetic Psychology

Founded by G. Stanley Hall in 1891; reorganized by Carl Murchison in 1925

Quarterly, beginning with March. Devoted to developmental, comparative, and clinical psychology, with briefly reported Replications and Refinements and occasional book reviews. Manuscripts may be sent to any member of the Editorial Board or directly to The Journal Press. Subscriptions should be sent directly to The Journal Press, and business communications addressed to the Managing Editor. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1891 at $15.00 per volume, plus transportation.

Journal of General Psychology

Founded in 1927 by E. B. Titchener and Carl Murchison

Quarterly, beginning with January. Devoted to experimental, physiological, and theoretical psychology, with briefly reported Replications, Refinements, and Comments. Manuscripts may be sent to any member of the Editorial Board or directly to The Journal Press. Subscriptions should be sent directly to The Journal Press, and business communications addressed to the Managing Editor. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1928 at $15.00 per volume, plus transportation.

Journal of Social Psychology

Founded in 1929 by John Dewey and Carl Murchison

Bimonthly, beginning with February. Devoted to studies of persons in group settings and of culture and personality; special attention to cross-cultural articles and notes, and to briefly reported Replications and Refinements. Manuscripts may be sent to any member of the Editorial Board or directly to The Journal Press. Subscriptions should be sent directly to The Journal Press, and business communications addressed to the Managing Editor. Three volumes annually. Per annum, $30.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1930 at $15.00 per volume, plus transportation.

Each author will receive 100 unbound reprints gratis. Additional reprints may be purchased at a nominal cost when the author returns page-proof.

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts 02657
U. S. A.

Club Rate for Five International Psychological Journals

	PER ANNUM
THE JOURNAL OF GENETIC PSYCHOLOGY (Quarterly, two volumes per year)	$20.00
GENETIC PSYCHOLOGY MONOGRAPHS (Quarterly, two volumes per year)	$20.00
THE JOURNAL OF GENERAL PSYCHOLOGY (Quarterly, two volumes per year)	$20.00
THE JOURNAL OF SOCIAL PSYCHOLOGY (Bimonthly, three volumes per year)	$30.00
THE JOURNAL OF PSYCHOLOGY (Bimonthly, three volumes per year)	$30.00
TOTAL	$120.00
TEACHERS AND STUDENTS EDITION, *Restricted Sale*	$50.00

The *Teachers and Students Edition* is open to any individual psychologist, teacher, student, or physician but requires an agreement that the edition will not be offered for sale till after five years from publication and an agreement that the edition will not be substituted for an institutional subscription.

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts 02657
U. S. A.

Genetic Psychology Monographs

Founded in 1925 by Carl Murchison

Quarterly, beginning with February. Devoted to developmental, comparative, and clinical psychology. Each number one or more complete researches. Manuscripts may be sent to any member of the Editorial Board or directly to The Journal Press. Subscriptions should be sent directly to The Journal Press, and business communications addressed to the Managing Editor. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1926 at $15.00 per volume, plus transportation.

NOTICE. Effective immediately we offer prior publication for certain selected monographs. Charges will be based on the announced rates for *The Journal of Psychology. For a limited period of time our policy is changed for regular monographs: Authors will be charged for cuts and expensive tables, but no page charge will be imposed.* Two hundred reprints of each monograph will be supplied gratis.

The listing of monographs for 1926 through 1963 will be printed in the first half of each volume of all five journals. Monographs published after 1963 will be listed in the second half of each volume.

Genetic Psychology Monographs *(continued)*

VOLUME 69—January-June, 1964

1. Occupational and industrial therapy in the treatment of psychiatric patients: A controlled study of efficacy in an intensive-treatment institution—M. Hundziak and B. Pasamanick
An analysis of Carl R. Rogers' theory of personality—M. S. Krause
Personality aspects of teaching: A predictive study—H. Zimiles, B. Biber, W. Rabinowitz, and L. Hay
Situational and personal antecedents of incipient alienation: An exploratory study—W. H. Jarrett and A. O. Haller
2. Age changes in children's mosaic responses from five to ten years—L. B. Ames and F. L. Ilg
A study of the interview process: The caseworker-client relationship—H. J. Meyer, E. F. Borgatta, and D. Fanshel
A study of introversion-extroversion as a personality dimension—R. M. Allen, H. M. Richer, and R. J. Plotnik
Psychological testing with Northwest Coast Alaskan Eskimos—C. E. Preston
A study in family pathology—M. B. Hans

VOLUME 70—July-December, 1964

1. Structure and function—H. Gutman
The Lowenfeld Mosaic Test: Norms for five-to-ten-year-old American public-school children and comparative study of three groups—L. B. Ames, F. L. Ilg, and J. August
Social and personality correlates of children's estimates of height—J. P. Shaffer
The development of ethnic attitudes in New Zealand school children—G. M. Vaughan
2. The antecedents of two varieties of vocational interests—L. E. Tyler
Personality in old age—P. E. Slater and H. A. Scarr
Variables affecting experimenter bias in a group situation—R. Rosenthal, C. J. Friedman, C. A. Johnson, K. Fode, T. Schill, R. C. White, and L. L. Vikan
Ordinal position and sex role identification—B. G. Rosenberg and B. Sutton-Smith
Personality correlates of accuracy of self-appraisal in four-year-old children—Paul S. Weiner

VOLUME 71—January-June, 1965

1. Author index for Genetic Psychology Monographs 1926-1964—L. W. Ferguson
A theoretical, developmental, and clinical investigation of embarrassment—J. M. Sattler
School readiness as evaluated by Gesell developmental, visual, and projective tests—F. L. Ilg, L. B. Ames, and R. J. Apell
Rorschach scales of regression in psychosis—S. Neiger, G. Alan, and D. A. Quirk
The effects of extra tactile stimulation on a group of institutionalized infants—L. Casler
Human eyelid conditioning: The recent experimental literature—L. E. Ross and T. F. Hartman
2. Social status and schoolroom aggression in third-grade children—R. Toigo
Overt and covert homosexual problems among male college students—L. J. Braaten and C. D. Darling
Role conceptions of the psychiatric aide—M. P. Lawton and A. E. Goldman
Changes in study of values scores during the four years of college—C. W. Huntley

VOLUME 72—July-December, 1965

1. A study of process and method in three-year-old twins and triplets learning to read—W. Fowler
Individualized teaching with therapeutic aims: A methodological study—M. Kohn and M. Rudnick
Ability and social adjustment at midlife of persons earlier judged mentally deficient—E. L. Miller
2. Development of incentive values in childhood—S. L. Witryol, D. J. Tyrrell, and L. M. Lowden
Longitudinal changes in intolerance and authoritarianism for subjects differing in amount of college education over four years—W. T. Plant
An NDEA counseling and guidance institute: Prediction, performance, and follow-up—R. H. Dana
The influence of language on classification: A theoretical model applied to normal, retarded, and deaf children—H. G. Furth and N. A. Milgram
Time estimates as affected by incentive class and motivational level—M. Felix
Clinical psychology: An historical survey—V. S. Sexton

VOLUME 73—January-June, 1966

1. Vulnerability in infants and young children: A pilot study—G. Heider
2. Racial preference and ego-identity among white and Bantu children in the republic of South Africa—A. J. Gregor and D. A. McPherson
Mother-child interaction in cerebral palsy: Environmental and psychosocial obstacles to cognitive development—E. Shere and R. Kastenbaum
Children's stories—L. B. Ames

VOLUME 74—July-December, 1966

1. Changes in personality for groups completing different amounts of college over two years—W. T. Plant and C. W. Telford
Brain dynamism in perception of reversible figures—G. K. Yacorzynski
Changes in Rorschach response throughout the human life span—L. B. Ames
Mnemonic virtuosity: A study of chess players—A. Binet (translated by M. L. Simmel and S. B. Barron)
2. The development of behavior patterns in children—B. B. Tyler, F. B. Tyler and J. E. Rafferty
Correlates of classroom functioning—L. M. Smith and B. B. Hudgins
Children's behavior syndromes and parental responses—R. L. Jenkins, E. NurEddin, and I. Shapiro

VOLUME 75—January-June, 1967

1. Studies of an elementary game of strategy—B. Sutton-Smith and J. M. Roberts
Child care practices anteceding three patterns of preschool behavior—D. Baumrind
The impact of cognitive maturity on the development of sex-role attitudes in the years 4 to 8—L. Kohlberg and E. Zigler
2. Differential factor structures in mathematical ability—P. S. Very
Perception of the vertical as a function of affective stimuli and extraneous stimulation in schizophrenic and normal subjects—J. O. Wells and W. E. Caldwell
Mid-life attainment of the mentally retarded: A longitudinal study—W. R. Baller, D. C. Charles, and E. L. Miller

VOLUME 76—July-December, 1967

1. Scalogram analysis of some developmental sequences of infant behavior as measured by the Bayley Infant Scale of Mental Development—R. Kohen-Raz
The leader and the lost: A case study of indigenous leadership in a poverty program Community Action Committee—L. A. Zurcher
A developmental study of the perception of people—R. N. Wolman
Human fears: A factor analytic study of three age levels—G. W. Russell
2. Correlates of voluntary social participation—J. E. Teele
Personality characteristics and developmental history of creative college women—R. Helson
A systematic study of therapeutic interventions in nursing homes—B. A. Stotsky

Genetic Psychology Monographs *(continued)*

VOLUME 76—July-December, 1967

1. Scalogram analysis of some developmental sequences of infant behavior as measured by the Bayley Infant Scale of Mental Development—R. Kohen-Raz
   The leader and the lost: A case study of indigenous leadership in a poverty program community action committee—L. A. Zurcher
   A developmental study of the perception of people—R. N. Wolman
   Human fears: A factor analytic study of three age levels—G. W. Russell
2. Correlates of voluntary social participation—J. E. Teele
   Personality characteristics and developmental history of creative college women—R. Helson
   A systematic study of therapeutic interventions in nursing homes—B. A. Stotsky

VOLUME 77—January-June, 1968

1. The fulfillment of promise: 40-year follow-up of the Terman Gifted Group—M. H. Oden
   A scale for parents' ratings: Some ipsative and normative correlations—R. N. Walker
   A methodological investigation of Piaget's concept of conservation of substance—A. C. Bittner and M. M. Shinedling
2. Sex differences in mental and behavioral traits—J. E. Garai and A. Scheinfeld

VOLUME 78—July-December, 1968

1. A theoretical and developmental investigation of delayed speech feedback—J. E. Waters
   Growth changes in factorial structure by age and sex—N. W. Dye and P. S. Very
   A multiaspect investigation of masculine development in kindergarten age boys—H. B. Biller
2. The optimum limits of testing on the Wechsler Intelligence Scales—M. Y. Quereshi
   A developmental study of the ego-ideal—L. Van den Daele
   The relationship between physical condition immediately after birth and mental and motor performance at age four—Nancy Edwards

VOLUME 79—January-June, 1969

1. Adolescents: From the views of the members of an informal adolescent group—T. P. Joseph
   Acculturation and childhood accidents among Chinese and Japanese Americans—M. Kurokawa
2. Dimensions of achievement-related behavior among lower-class Negro parents—D. Solomon, R. J. Parelius, and T. V. Busse
   Dimensions in the judgment of illness—J. S. Kerrick
   Reversal shifts and problem-solving abilities at ages 4-7—S. B. Kastner
   The judgment of body landmark heights—H. Nash
   Recognition of body surface regions—H. Nash

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts 02657
U. S. A.

PUBLICATION IN THE JOURNAL OF PSYCHOLOGY

The Journal of Psychology provides a unique service to authors. Articles accepted for publication in this periodical appear in print with the least possible amount of delay. In line with this policy, all steps of the acceptance procedure are maximally streamlined, and the author's work is presented with a bare minimum of editing or restriction of any kind. *The Journal of Psychology* is the ideal medium for presentation of timely professional material.

Authors are requested to exercise unusual care in the preparation of papers for this journal. Accuracy, grammar, and style should be checked and rechecked before a clean version of the manuscript is submitted. Careful attention should be paid to the General Instructions on the back cover of each issue and to the announcement on the inside front cover.

Authors are particularly requested to note the fact that all parts of the manuscript—including references, footnotes, and tables—must be double-spaced.

$30.00 per annum
$15.00 per volume
$7.50 per half volume

BIMONTHLY
Three volumes per year
Immediate publication

July, 1969
Volume 72, Second Half

(Founded in 1935 by Carl Murchison)

The Journal of PSYCHOLOGY

The General Field of Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold

under the restrictions of a bilateral contract, and may not be resold for a period of five years from the date of publication.

JULY, 1969

(Manuscripts are printed in the order of final acceptance)

Approximate randomization tests 143
By Eugene S. Edgington

Who seeks treatment? A study of psychiatrically disabled veterans . . . 151
By Herbert C. Archibald, David E. Parker, M. Kent Smith, and Read D. Tuddenham

A comparison of "real" *vs.* "ideal" self with a self-actualization inventory . 159
By John R. Braun and Patricia Asta

The clinical utility of Block's ego-resiliency and ego-control scales . . . 165
By Dennis Kneff, Robert T. Bodensteiner, Thomas W. Vodde, and Malcolm Gynther

Factorial study and analysis of factor variance of general adaptability self-ratings from a combined sample of male and female subjects . . . 169
By Rio Sciortino

(over)

Copyright © 1969 by The Journal Press
Provincetown, Massachusetts, U. S. A.
All rights reserved

Published at 35 New St., Worcester, Mass.
Second-class postage paid at Worcester, Mass.
Send all mail to Managing Editor, 2 Commercial Street, Provincetown, Massachusetts 02657

The fear of death and the fear of dying 179
By Lora-Jean Collett and David Lester

Social role perceptions of children in Hutterite communal society 183
By Shirin Schludermann and Eduard Schludermann

Peer influences on preschool children's willingness to try difficult tasks . . 189
By E. Paul Torrance

The use of matrix rather than serial presentations of nonsense syllables . . 195
By Alan Krichev and Brenton Hazlewood

Interstressor consistency of subject differences in autonomic activity . . 199
By Gary W. R. Patton

Magnitude scaling of foods 211
By Judith W. Onley, Nina Sullivan, and Susan H. Borcyck

The color-word interference test: The effects of varied color-word combinations upon verbal response latency 219
By Joyce L. Sichel and Kenneth A. Chandler

The self-concept of creative adolescents 233
By Charles E. Schaefer

Developmental study of social role perception among Hutterite adolescents . 243
By Shirin Schludermann and Eduard Schludermann

Raven's Progressive Matrices: More on construct validity 247
By Henry R. Burke and William C. Bingham

Scale checking style as a function of age and sex in Indian and Hutterite children 253
By Shirin Schludermann and Eduard Schludermann

A comparison of volunteers and nonvolunteers for astronaut training . . . 263
By L. K. Waters and Rosalie K. Ambler

Quantitative behavior of warring nations 269
By John Voevodsky

Published as a separate and in *The Journal of Psychology,* 1969, **72**, 143-149.

APPROXIMATE RANDOMIZATION TESTS*

Department of Psychology, University of Calgary

EUGENE S. EDGINGTON

A. INTRODUCTION

In an article on randomization tests, approximate randomization tests were briefly discussed (1). This paper will consider approximate randomization tests in further detail. First, however, it is necessary to correct a statement in the earlier article. In a randomization test, the significance of an obtained statistic is the proportion of the statistics in the sampling distribution that are *as large as* the obtained statistic, *not,* as was erroneously stated, the proportion that exceed the obtained statistic. The correction is required for the computed significance value to be the probability, under the null hypothesis, of obtaining such extreme results.

B. TERMINOLOGY

The term *entire sampling distribution* will be used to refer to the sampling distribution of a randomization test consisting of all divisions or pairings of measurements that are equally probable under the null hypothesis and random assignment, or to refer to the distribution of statistics associated with the distribution of divisions or pairings. For example, given the null hypothesis of no difference between the effects of Treatments A and B and random assignment of 10 subjects to A and five to B, the entire sampling distribution of divisions consists of every possible division of the 15 obtained measurements that assigns 10 to Treatment A and five to Treatment B. If one were to compute a difference between means for each division in the distribution, the distribution of differences between means thus obtained would be the entire sampling distribution of differences between means.

An *approximate sampling distribution* is a smaller distribution of pairings or divisions of measurements (or a distribution of statistics associated with them) obtained by taking a random sample of all of the pairings or divisions in the entire sampling distribution.

The term *approximate randomization test* refers to a randomization test that uses an approximate sampling distribution instead of the entire sampling distribution in determining the significance of an obtained statistic.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on March 31, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

C. Purpose of Approximate Randomization Tests

The advantages of randomization tests over parametric and rank order tests have been widely recognized. The only serious objection that has been raised against the use of randomization tests is that the amount of computation required is excessive. The purpose of approximate randomization tests is to reduce the amount of computation for a randomization test to a practical level, thereby overcoming this objection. The following example will illustrate the reduction in computation that can result from using an approximate randomization test.

Suppose that in an experiment 13 subjects are randomly assigned to 13 different magnitudes of an independent variable and one measurement is obtained from each subject. The null hypothesis is that variation in magnitude of the independent variable has no effect on the dependent variable; i.e., at the time he was measured every subject gave the same response that he would have given at that time to any of the other 12 magnitudes of the independent variable. This hypothesis is tested by computing a product-moment correlation coefficient or some other correlation statistic. Under the null hypothesis and random assignment there are 13! (factorial 13) = 6,227,020,800 equally probable pairings of the 13 magnitudes of the independent variable with the 13 response magnitudes. Thus, to determine with a randomization test what proportion of the sampling distribution of pairings gives as large a correlation statistic as the obtained value would require the computation of over six billion correlation statistics. Even an electronic computer that paired the measurements and computed a correlation statistic every second would take 197 years of continuous, 24-hours-a-day operation to compute all of the correlation statistics in the entire sampling distribution.

Now consider the application of an approximate randomization test to the same situation. An electronic computer is programmed to use a table of random numbers to simulate taking a random sample, *with replacement,* of 999 pairings from the entire sampling distribution and then to compute a correlation statistic for each of the selected pairings. This provides an approximate sampling distribution of 999 correlation statistics within minutes, in contrast to the 197 years required for the entire sampling distribution.

D. Determination of Significance for Approximate Randomization Tests

Now consider the procedure for determining the significance of results when approximate randomization tests are performed. The approximate sampling distribution is *not* used in exactly the same way as the entire sampling distribution in determining significance. Instead of determining the proportion of the statistics in the approximate sampling distribution that are as large as the obtained statistic, one determines the proportion within the joint distribution consisting of the approximate sampling distribution *plus* the obtained statistic that are as large as the obtained statistic. For example, when the approximate sampling distribution consists of 999 randomly selected statistics, as in the correlation problem, one determines how many of the 1000 correlation statistics (the 999 in the approximate sampling distribution plus the obtained statistic) are as large as the obtained statistic. (Since the obtained statistic is as large as itself, there can be no less than one of the 1000 statistics as large as the obtained statistic.) This number is divided by 1000 to get the probability under the null hypothesis of getting such a large correlation statistic as the one obtained. One can then use this probability to decide whether to reject the null hypothesis. For instance, with a significance level of .05 the null hypothesis is rejected when the computed probability is no greater than .05; and with a significance level of .01, the null hypothesis is rejected when the computed probability is no greater than .01. That is, the null hypothesis is rejected at the .05 level whenever the obtained statistic falls within the upper five percent of the 1000 statistics (including the obtained statistic itself) and at the .01 level whenever the obtained statistic falls within the upper one percent of the 1000 statistics.

It will now be shown that this procedure is valid: i.e., that when the null hypothesis is true the probability of rejecting the null hypothesis is equal to the level of significance. For example, when the null hypothesis is true, for this procedure to be valid the probability of rejecting at the .01 level must be .01.

Random assignment in conjunction with the null hypothesis permits one to regard the obtained statistic as a random sample of one statistic taken from the entire sampling distribution of statistics. When the obtained statistic is added to the 999 randomly selected statistics there are, under the null hypothesis, 1000 statistics randomly selected from the entire sampling distribution. The obtained statistic can be regarded as the first statistic to be drawn. In random sampling there is, of course, no systematic tendency for the first drawn value to be either larger or smaller than values drawn later. So, for 1000 randomly selected statistics the probability that the obtained statistic (the first drawn statistic) will be within the upper half of the statistics is ½, that it will be within the upper quarter is ¼, and so on. Thus the probability is .05 that, when the null hypothesis is true, the obtained statistic will fall within the upper 5 percent of the 1000 statistics, and the probability is .01 that it will fall within the upper one percent of the 1000 statistics. Now, the recommended procedure rejects at the .05 level when the obtained statistic is in the upper five percent of the statistics and at the .01 level when it is in the upper one percent; so the probability under the null hypothesis of rejecting at the .05 level is .05 and at the .01 level is .01. Approximate randomization tests are therefore *valid, not approximately valid,* procedures for testing null hypotheses. Probabilities computed from approximate sampling distributions are not simply approximations to probabilities based on entire sampling distributions, but are probabilities in their own right.

The same argument could have been used to show the validity of the proposed procedure if sampling *without* replacement had been carried out, but the following discussion of the power of the approximate randomization test will be considerably simplified by restricting it to sampling distributions obtained by sampling *with* replacement.

E. Power of Approximate Randomization Tests

The next consideration is the relative power of the approximate randomization test based on an approximate sampling distribution of 999 statistics. First the following statement will be proved:

Statement 1: The probability is .99 that no more than 17 statistics from the upper one percent of the entire sampling distribution will be in the approximate sampling distribution.

Then it will be shown that Statement 1 implies Statement 2:

Statement 2: The probability is .99 that, whenever use of the entire sampling distribution would lead to judging an obtained statistic to be significant at the .01 level, use of the approximate sampling distribution would lead to the assignment of a probability no greater than .018.

For the proof of Statement 1 the entire sampling distribution is to be regarded as consisting of two kinds of statistics: those that are in the upper one percent of the distribution and those that are not. When the entire sampling distribution is randomly sampled with replacement, for a single draw there is a probability, $p = .01$, of drawing a statistic from the upper one percent of the distribution and a probability $q = .99$ of not doing so: that is, of drawing a statistic from the lower 99 percent of the distribution. Replacement after each draw ensures that p and q stay constant over all draws. The binomial distribution of the number of statistics from the upper 1 percent of the entire sampling distribution for $N = 999$, $p = .01$, and $q = .99$ is almost a normal distribution with $\mu = Np = 9.99$ and $\sigma =$

$\sqrt{Npq} = 3.14$. Ninety-nine percent of the area under a normal curve lies below the point that is 2.33 standard deviations above the mean. Consequently the probability is .99 that the number of statistics from the upper one percent of the entire sampling distribution contained in the approximate sampling distribution is no greater than $9.99 + (2.33)(3.14) = 17.31$. This verifies Statement 1, since the discrete frequency 17 is assumed to occupy the interval 16.5 to 17.5 in a normal distribution.

Statement 1 also has been verified by computations by an electronic computer on the exact binomial distribution (not the normal curve approximation).

To see that Statement 1 implies Statement 2, it is necessary to appreciate that Statement 1 is true regardless of the magnitude of the obtained statistic, inasmuch as the selection of the statistics in the approximate sampling distribution is in no way affected by the magnitude of the obtained statistic. Therefore Statement 1 is true even when the obtained statistic is large enough to be judged significant at the .01 level by means of the entire sampling distribution. This fact enables one to derive Statement 2 from Statement 1. An obtained statistic would be judged significant at the .01 level by means of the entire sampling distribution only if it had a value that lies within the upper one percent of the entire sampling distribution of statistics. The only statistics in the approximate sampling distribution that might possibly be as large as the obtained statistic, therefore, would be those from the upper one percent of the entire sampling distribution. This fact, in conjunction with Statement 1 (which has already been verified), leads to this conclusion: When an obtained statistic would have been judged significant at the .01 level by means of the entire sampling distribution, the probability is .99 that there will be no more than 17 values in the approximate sampling distribution that are as large as the obtained statistic. When no more than 17 statistics in the approximate sampling distribution of 999 statistics are from the upper one percent of the entire sampling distribution, no more than *18* (17 in the approximate sampling distribution plus the obtained statistic itself) of the 1000 will be as large as the obtained statistic. Statement 2, therefore, is true.

Statement 2 is correct whether or not the null hypothesis is true: i.e., whether there is a treatment effect or not. Statement 2 is of special interest however when the null hypothesis is false because it shows that an approximate randomization test based on an approximate sampling distribution of 999 statistics is almost as powerful as a randomization test employing the entire sampling distribution.

Analogous computations and arguments in regard to the binomial distribution of the number of statistics from the upper five percent of the entire sampling distribution for $N = 999$, $p = .05$, and $q = .95$ lead to this statement:

Statement 3: The probability is .99 that whenever use of the entire sampling distribution would lead to judging an obtained statistic to be significant at the .05 level, use of the approximate sampling distribution would lead to the assignment of a probability no greater than .066.

Like Statement 2, Statement 3 has also been verified by electronic computer computations on the exact binomial distribution.

One can, of course, follow the same sort of procedure to determine probability statements comparable to Statements 2 and 3 for other levels of significance or for approximate sampling distributions of other sizes. An approximate sampling distribution with more than 999 statistics has more power, but it can be seen that this number provides sufficient power for many purposes.

F. SUMMARY

An approximate randomization test is a randomization test in which the significance of an obtained statistic is determined by using an approximate sampling distribution consisting of a random sample of the statistics in the entire sampling distribution.

The purpose of approximate randomization tests is to reduce the amount of computation for a randomization test to a practical level, thereby overcoming the only serious objection that has been raised against the use of randomization tests.

In an approximate randomization test the significance of an obtained statistic (such as a correlation coefficient or a difference between means) is determined by reference to a distribution composed of the approximate sampling distribution *plus* the obtained statistic. The significance is the proportion of statistics within this distribution that are as large as the obtained statistic. This proportion is the probability under the null hypothesis of getting such a large statistic as was obtained. This procedure is not just a procedure for providing estimates of the probabilities that would be obtained by using the entire sampling distribution, but is valid in its own right: when the null hypothesis is true the probability of rejecting the null hypothesis is equal to the level of significance used.

When the approximate sampling distribution consists of 999 statistics, the power of the approximate randomization test is almost equal to that of the

randomization test employing the entire sampling distribution. Specifically, the probability is .99 that whenever use of the entire sampling distribution would lead to judging an obtained statistic to be significant at the .01 level, use of the approximate sampling distribution would lead to the assignment of a probability no greater than .018. Also, the probability is .99 that whenever use of the entire sampling distribution would lead to judging an obtained statistic to be significant at the .05 level, use of the approximate sampling distribution would lead to the assignment of a probability no greater than .066. These probability statements are true for approximate sampling distributions of 999 statistics selected at random, with replacement, from the entire sampling distribution regardless of the size of the entire sampling distribution.

REFERENCE

1. EDGINGTON, E. S. Randomization tests. *J. of Psychol.*, 1964, **57**, 445-449.

Department of Psychology
University of Calgary
Calgary, Alberta
Canada

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 151-157.

WHO SEEKS TREATMENT? A STUDY OF PSYCHIATRICALLY DISABLED VETERANS*

Mental Hygiene Clinic, Oakland VA Office[1]

HERBERT C. ARCHIBALD, DAVID E. PARKER, M. KENT SMITH,
AND READ D. TUDDENHAM

A. INTRODUCTION

Every public facility for treatment is concerned with the degree to which it serves its potential population of eligible patients. This is particularly true of the Veterans Administration because the nation has long since committed itself to providing optimal care to all former members of the armed forces who meet eligibility requirements. The present study compares two groups of psychiatrically disabled veterans: those who receive psychiatric treatment in an outpatient Veterans Administration Mental Hygiene Clinic; and those who do not receive treatment, although equally eligible. It is part of a continuing research investigation both of the treatment process and of the characteristics of the outpatient psychiatric veteran population. However, the present concern is not with treatment, but rather with those factors associated with the veteran's decision to enter psychotherapy.

1. *Previous Studies*

In 1962, DeBakey and Beebe (3) reviewed medical follow-up studies with the veteran population, which were then either completed or in progress. At that time, the bibliography included some 99 reports covering the entire field of medicine. Only a fraction of the studies were in the general field of psychiatry. One such study, that by Sullivan (6) in 1955, is of interest because it, like the present investigation, was conducted at the Oakland, California, Veterans Administration Clinic. Sullivan compared treatment and nontreatment groups with respect to the MMPI. His nontreatment group consisted of veterans with a primarily psychiatric diagnosis who came to the outpatient clinic for a periodic evaluation of their compensation status, but who had not been treated there. The treatment group consisted of active pa-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on March 31, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Attached to S.F. VA Hospital and the University of California (R.D.T.), Berkeley, California.

tients at the Veterans Administration Clinic. Differences between the groups were not great, but those who came for treatment were the more disturbed. It proved possible to create an MMPI scale which differentiated patients from nonpatients with a 72 percent hit rate on cross-validation. Items chosen more often by patients related to excessive morality, excessive conscientiousness, guilt, and self-defeating neurotic trends.

In 1958, Sullivan, Miller, and Smelser (7) did a related study at the same clinic on factors in "length of stay" and "progress in psychotherapy." Only age, education rating, and occupational rating were reliable predictors of length of stay. They reported that on other variables the results were extremely inconsistent from sample to sample. Improvement, as rated by staff members, was associated with higher occupational levels and with less serious psychopathology.

Diener and Young (4) in 1961 studied psychiatrically diagnosed veterans in Colorado, and failed to find significant relationships between coming or not coming for treatment on the one hand, and on the other hand, degree of disability, number of compensation cuts, or husband *vs.* son role played by veterans in their families. However, extending the sample to the entire population of such cases in Colorado, in 1964, Dworin, Green, and Young (5) found that both degree of disability and nearness to the mental hygiene clinic had significant correlations (.33 and .37 respectively) with coming to the clinic for treatment.

2. *Present Study*

The concern of the Oakland Veterans Administration Clinic to check its own effectiveness, especially in the light of inconsistencies in previous investigations, suggested the desirability of further research. Specifically, the clinic staff wished to extend Sullivan's MMPI study to explore differences between treated and nontreated veterans in respect to overt symptoms. Further, the clinic wished to evaluate the implications of the Dworin *et al.* (5) and Sullivan *et al.* (7) papers that demographic and geographical factors may be important in seeking treatment.

B. METHOD

1. *Sample*

The "treatment" group consisted of 67 psychiatrically disabled veterans in active psychotherapy. The length of treatment was from one to over 100 interviews. All but 10 cases had had more than nine interviews, the median length of stay in this clinic (7).

The "nontreatment" group consisted of 68 veterans who came to the clinic only for a review of their neuropsychiatric compensation claims. Successive neuropsychiatric re-evaluation cases were referred to the mental hygiene clinic until the group approximated the treatment group in size. These veterans had never sought outpatient psychiatric treatment in the Veterans Administration.

2. *Procedure*

Patients in both treatment and nontreatment groups were asked to complete a questionnaire[2] containing a checklist of somatic symptoms of stress, and items concerning subjective feelings of discomfort, personal worries, phobic fears, and anxiety. This questionnaire also included sociological items pertaining to job stability, marital status, and history, etc.

The nontreatment cases were interviewed by a staff member, after they had completed the questionnaire, to check further upon whether or not they had had any psychotherapy, and if so, its nature and amount. On the basis of this interview, veterans were rated on 12 scales, each consisting of five carefully defined steps. Variables included were speech quality, motor behavior characteristics, anxiety level, depression, hypomania, anger, hostility, abnormal mental trends, attitude toward treatment, frankness, emotional distance, and dependence.

C. RESULTS

Fifty-seven items of the questionnaire could be described as "complaints" to be checked or not checked. The questionnaire included 73 items for combat veterans (10 percent of the clinic population). Eight questions were deleted for those who were service-connected but saw no combat. Twenty-two symptoms were checked in each group for "in service," "at discharge," and for the present (time of interview). Demographic facts and a complete neuropsychiatric examination were available in every case.

Notwithstanding the fact that the checklists were filled on the day of compensation examination by the untreated veterans but after an average of seven hours of psychotherapy for the treated veterans, no change was noted in the latter's basic attitude toward symptoms or the preponderance of psychological symptoms.

Included were such symptoms as severe headache, or irritability, fears—such as fear of jet plane noises—and "worries"—for example, about letting others down, being injured, or becoming insane. Of these 57 symptoms, 44 were checked more frequently by the patients than by the nonpatients; only 13 more frequently by nonpatients than by patients. This ratio is significant beyond the .001 level. Clearly, members of the treatment group either experienced more symptoms than members of the nontreatment group or were more willing to acknowledge having them. Within the nontreatment group, interview ratings correlated meaningfully with questionnaire items, suggesting that those who checked the most symptoms were seen as being, for example, more "anxious" and "depressed" by an observer.[3]

---

[2] This questionnaire is described at greater length in Archibald *et al.* (1, 2). Many of its items clearly differentiated combat stress patients both from healthy combat veterans and from psychiatrically disturbed noncombat veterans.

Clinical experience had led the authors to suspect that patients in treatment are relatively more likely to acknowledge phobic fears and worries; whereas the nontreatment cases are relatively more likely to report symptoms having a strong somatic component. To test this hypothesis, the two groups were compared with respect to choice frequency (i.e., frequency of item checking) summed across the 10 most clearly somatic items *versus* choice frequency summed across the 27 more clearly "psychological items," consisting largely of worries and fears. The χ^2 was significant (one tail) at the five percent level. While less certain than the finding as to overall number of symptoms in the two groups, there is a strong indication that patients view their difficulties as more psychological in nature than do nonpatients; whereas the nonpatients emphasize their somatic symptoms.

The specific symptom items differentiating the two groups at various significance levels are listed in Table 1. The psychological aspects of the concerns of patients in treatment are apparent in the table. While the nontreatment group checked somatic symptoms more often than did the treatment group, only one—"sweaty hands or feet"—reached statistical significance. A few other items approached the 10 percent level. More nontreatment than treatment cases denied sexual difficulties, and more nontreatment were still active in a veterans' organization. Asked to compare their general health at present to their health status while in service, nonpatients had a tendency to use the more extreme categories, "better" and "much worse," while patients more often saw themselves as unchanged. Replication of the study would be required to establish the meaningfulness of these last differences.

Sociological variables, such as marital status and job stability, were not differentiating; but social status as measured by educational level and occupational level was a factor with regard to staying in therapy (Table 2).

---

[3] Since the interviewer had access to the questionnaires to check for omitted items or skipped pages, there is a possibility that his ratings may have been contaminated by the subjects' responses to the questionnaire, although the ratings were intended to be independent.

TABLE 1
DICHOTOMOUS QUESTIONNAIRE ITEMS WHICH SIGNIFICANTLY DIFFERENTIATE TREATMENT *vs.* NONTREATMENT GROUPS

Significant at 1% level	Significant at 5% level	Significant at 10% level
	A. *Items checked more often by the treatment group*	
Irritability	Depression	Sighing and yawning (Hyperventilation)
Alcohol to excess	Worry about insanity	Other things (than therapy) have helped your condition
Worry about specific acts you were ashamed of	Worry about fear of hurting others	
Matters overlooked by clinic which would help in understanding your condition	Additional symptoms (added by veteran)	
	B. *Items checked more often by the nontreatment group*	
Sweaty hands or feet	(none)	(none)

TABLE 2
COMPARISON OF EDUCATIONAL AND OCCUPATIONAL LEVELS

Variable	Stayers		Nonstayers		Nontreated	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Educ. rating	2.84	1.19	3.42	1.26	3.61	1.47
Occup. rating	4.27	1.87	4.90	1.82	4.91	2.10

Note: Ratings on the Warner socioeconomic scales for Educational and Occupational level indicate that there is a trend for lower educational and occupational level to be associated with not seeking treatment or seeking but not staying in treatment. The differences between the Stay and Nonstay group on Educational and Occupational are significant at the .001 level and the .01 level respectively. While the nontreated educational and occupational levels are not significantly different from the nonstayers, their scores show a trend in the expected direction.

The accessibility of the clinic to the veteran's place of residence also seemed to be a factor.[4] In checking the travelling distance that these nontreatment subjects would have to come, it was discovered that many lived as far as 100 miles or more from the clinic. They could travel this distance for periodic compensation examinations, but could hardly have been expected to have become patients at this clinic. Yet, many come from large urban centers where state mental health services are available, and private psychiatric services can be obtained locally in many communities and paid for by the Veterans Administration. Many of these patients, then, could have obtained some form of treatment had they desired it. This consideration led us to suspect that many patients in the treatment group would not have obtained treatment if they had lived outside a reasonable commuting radius from the clinic. To check on the influence of distance *per se* on use of the clinic, the home town of each subject, treatment or nontreatment, was obtained and scored according to the number of 10-mile intervals between it and the clinic. The distances were measured from an ordinary road map of the Bay Area; they were taken "as the crow flies" and, hence, only approximated travelling time. Nonetheless, differences between the two groups were more marked than had been anticipated. Every member of the treatment group lived within a 30-mile radius of the clinic, while only 22 of 68 in the nontreatment category lived within this radius. The two distance distributions were compared by means of the Mann-Whitney U-test. The result was significant far beyond the one percent point ($= > 7$).

---

[4] The above finding suggested that distance from the clinic might also be importantly related to the number of treatment interviews. A sample of 100 cases who had terminated treatment were selected from the files, and the actual road distance of their home town from the clinic was obtained. This "mileage from clinic" was then correlated with the number of interviews prior to termination. The correlation was nonsignificant (.08) and remained so when calculated for just those 63 patients living outside Oakland. Thus, distance operated for patients in an all-or-nothing fashion. If they decided to become patients at all, distance was not a factor in determining how long they remained in treatment.

D. Conclusions

In this study of utilization of an outpatient Veterans Administration mental hygiene clinic facility, the most conspicuous difference between veterans who elected and those who did not elect psychiatric treatment is distance of residence from the clinic. However, the fact that nontreatment cases living at a distance from the clinic did *not* elect treatment at easily accessible state or county facilities or from psychiatrists in their own communities paid by the Veterans Administration suggests that the causal factors in accepting or avoiding treatment are more complicated than they appear. Those who sought and stayed in treatment tended to be of higher educational and occupational level. Study of the reasons for nonutilization or underutilization of the available non-Veterans Administration facilities by veterans appears to be indicated.

Patients and nonpatients appear to differ also in the way they formulate their difficulties. Patients reported more symptoms and their complaints were more psychological in nature. Nonpatients emphasized somatic symptoms, and indeed often denied psychological concerns. Comments, such as the following, were typical in this group:

"I don't understand why I was sent up here. Does that mean someone thought I was emotionally unstable?"

"I don't understand why they call me a nervous case. I had a back injury."

"I've never been emotional in my life. I don't worry about anything. I have a wonderful wife."

However, it is uncertain whether the patients elected psychiatric treatment because they were more sensitive to the psychological component of their difficulties or whether, prompted by unknown factors, they came in for treatment and, in the course of it, learned to see their problems in a psychological light.

REFERENCES

1. ARCHIBALD, H. C., & TUDDENHAM, R. D. Persistent stress reaction after combat—a 20-year follow-up. *Arch. Gen. Psychiat.*, 1965, **12**, 475-481.
2. ARCHIBALD, H. C., LONG, D. M., MILLER, C., & TUDDENHAM, R. D. Gross stress reaction in combat—a 15-year follow-up. *Amer. J. Psychiat.*, 1962, **119**, 317-322.
3. DEBAKEY, M. E., & BEEBE, G. W. Medical follow-up on veterans. *Amer. J. Psychiat.*, 1963, **182**(11), 1103-1109.
4. DIENER, R. G., & YOUNG, H. H. Factors contributing to requests for mental hygiene clinic treatment by veterans with psychiatric disorders. *J. Clin. Psychol.*, 1961, **17**, 397-399.
5. DWORIN, J., GREEN, J. A., & YOUNG, H. H. A follow-up study of relationships between distance from the clinic, degree of disability, and requests for psychiatric treatment. *J. Clin. Psychol.*, 1964, **20**(3), 393-395.
6. SULLIVAN, P. L. A treatment evaluation program. Program guide, P & M Service, United States Veterans Administration Publications, G6M2, 1955. Part 10.
7. SULLIVAN, P. L., MILLER, C., & SMELSER, W. Factors in length of stay and progress in psychotherapy. *J. Consult. Psychol.*, 1958, **22**(1), 1-9.

Mental Hygiene Clinic Section
Veterans Administration VA Outpatient Clinic
1305 Franklin Street
Oakland, California 94612

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 159-164.

A COMPARISON OF "REAL" *VS.* "IDEAL" SELF WITH A SELF-ACTUALIZATION INVENTORY*[1, 2]

Department of Psychology, University of Bridgeport

JOHN R. BRAUN AND PATRICIA ASTA

A. INTRODUCTION

The Personal Orientation Inventory (POI) contains 150 paired-opposites items scored for 12 scales attempting to measure self-actualization (10). Individuals nominated as nonself-actualized have obtained significantly lower (less favorable) scores on 11 of these scales than those nominated as self-actualized (9, 10). There have been similar findings with "highs" *vs.* "lows" on the Neuroticism scale of the Eysenck Personality Inventory (6), alcoholics *vs.* normals (12), and beginning therapy patients *vs.* patients in therapy for two years (11). Ilardi and May (5) found that POI reliability compares favorably with that of the MMPI, and Grossack (4) has studied Edwards Personal Preference Schedule and 16 PF correlates of self-actualization scores. Damm (3) reported a significant positive correlation between performance on the POI and the Remote Associates Test (a measure of creativity), and Braun and Asta (1) have found significant positive correlations between the Original Thinking Scale of the Gordon Personal Inventory and five of the POI scales.

Braun and LaFaro (2) had four groups of undergraduates complete the POI under standard instructions, and later, under instructions to make either a good impression or appear well-adjusted. Results showed a *less* favorable score when faking in 45 of the 48 comparisons between standard and faked-administration means, with 23 of the 48 *t* tests significant at the .05 level or better. All 23 of the significant findings involved *lower* faked *versus* standard administration scores. It was concluded that "Apparently the values deemed to be those of self-actualizing persons are not those judged likely to create a good impression or to be indicators of good adjustment by the typical college student (or convicted felon!)."

---

* Received in the Editorial Office, Provincetown, Massachusetts, on April 1, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Presented at the annual convention of the Eastern Psychological Association, Washington, D.C., April, 1968.

[2] The authors are indebted to Mrs. Frances Brown for her kind cooperation in obtaining the data on Groups I and II.

Rapaport (7) studied the effects of real-self *vs.* ideal-self instructions on the MMPI performance of 48 military psychiatric patients. Ideal-self means were significantly more favorable than real-self means for the majority of the scales. The present study compared POI performance under real-self and ideal-self instructions. Persons rather consistently score lower on self-actualization when trying to make a good impression than they do when trying to be honest. How would they *like* to be in terms of self-actualization?

B. Method

Three groups of undergraduates were used as subjects. Group I consisted of students enrolled in evening sections of introductory psychology (real-self $N = 60$; ideal-self $N = 67$). Group II consisted of student nurses enrolled in introductory psychology (real-self $N = 51$; ideal-self $N = 52$). Group III consisted of volunteers recruited from introductory psychology (N for both real and ideal self $= 63$). There were approximately equal numbers of men and women in Groups I and III, while Group II had all women. Groups I and II were tested during regular class meetings, while Group III was tested outside of class.

All groups completed the POI on two occasions. The *N*s for the two administrations differ slightly for some groups due to absences from class. Groups I and II had ideal-self administration first, and approximately one month later, real-self administration. Group III had real-self administration first, and one week later, ideal-self.

Groups I and II did not put any identifying data on their answer sheets, while Group III used only self-selected pseudonyms to enable matching of first and second administration protocols for statistical purposes. Significance of differences between means for the two administrations was evaluated by t tests for Groups I and II, and by Sandler's A-statistic (8) for Group III.

Real-self instructions emphasized being honest and forthright "answer as you really are." Ideal-self instructions indicated ". . . we want you to answer so as to tell us about your ideal self. That is, what sort of person would you ideally like to be? Answer *not* in terms of your real self, the kind of person you are now; but rather, answer in line with the kind of person you would *like* to be: your ideal self."

C. Results and Discussion

Table 1 presents means, standard deviations, and results of significance tests for the two administrations. The direction of difference between real-self and ideal-self means was consistent across all three groups for 10 of the 12 POI scales, with 21 of the 36 *t* or A-tests significant at the .05 level or better. Time Competence, Self-Actualizing Value, Spontaneity, Self-Regard, Nature of Man, Constructive, and Synergy were increased in all three groups; while Existentiality, Feeling Reactivity, Self-Acceptance, and Acceptance of Aggression were decreased. For each of the above scales, the difference between real- and ideal-self means was significant in at least one, and more often, two or three, of the groups. The Inner Directed and Capacity for Intimate Contact scales were not affected consistently by the different instructions across all three groups, and for these scales, none of the comparisons between real- and ideal-self means reached significance.

All 12 POI scales are assumed to measure an aspect of self-actualization. Under ideal-self instructions, our subjects moved in the direction of greater actualization on six scales, while moving toward less actualization on four other scales. From scale descriptions in the POI manual (10) we would say that under ideal- *vs.* real-self instructions, they prefer to live in the present rather than past or future; hold, rather than reject, the values of self-actualizing people; freely express feelings behaviorally rather than being fearful of doing so; have high, rather than low, self-worth; see man as essentially good rather than evil; and see the opposites of life as meaningfully related rather than antagonistic. On the other hand, we would also say that they prefer to be rigid, rather than flexible, in application of values; insensitive, rather than sensitive, to own needs and feelings; unaccepting, rather than accepting, of the weaknesses of self; and denying, rather than accepting, of feelings of anger or aggression.

With six scales increased and four decreased under ideal-self instructions, there is no discernible tendency for subjects to wish to be either higher or lower in self-actualization than they think they already are. This conclusion is supported by the lack of any significant change on the Inner Directed scale, which contains 127 of the 150 items, and which is regarded by Knapp (6) as the single most representative overall measure of self-actualization.

It would be of interest to determine if the same trends found in the present study would be present with subjects of different age, education, socioeconomic status, etc. In addition, the contribution of social desirability to the various POI scales should be investigated. Are the items on the scales which were increased under ideal-self instructions significantly higher in mean social desirability rating than those which appear on the scales that were decreased under this instruction?

TABLE 1
MEANS AND STANDARD DEVIATIONS FOR "REAL" *versus* "IDEAL" SELF

			Group	
Parameter	Instruction	I	II	III
		Time Competence		
Mean	Real Self	15.47	15.13	15.40
SD	Real Self	3.31	3.91	3.34
Mean	Ideal Self	17.15**	16.79*	16.70**
SD	Ideal Self	2.96	2.84	2.98
		Inner Directed		
Mean	Real Self	80.93	78.75	78.08
SD	Real Self	11.52	11.69	10.35
Mean	Ideal Self	81.85	75.48	79.67
SD	Ideal Self	9.33	12.68	9.27
		Self-Actualizing Value		
Mean	Real Self	19.18	18.49	17.94
SD	Real Self	2.78	3.43	3.18
Mean	Ideal Self	20.12	19.65	20.11**
SD	Ideal Self	2.52	2.72	3.45
		Existentiality		
Mean	Real Self	19.53	18.27	19.08
SD	Real Self	4.06	4.55	3.50
Mean	Ideal Self	17.84*	15.79**	15.79**
SD	Ideal Self	3.88	3.64	4.45
		Feeling Reactivity		
Mean	Real Self	16.00	14.63	15.43
SD	Real Self	2.68	3.00	2.73
Mean	Ideal Self	14.82*	14.21	14.06**
SD	Ideal Self	2.68	2.80	2.95
		Spontaneity		
Mean	Real Self	11.45	10.60	11.00
SD	Real Self	2.35	2.82	2.48
Mean	Ideal Self	12.04	11.39	12.33**
SD	Ideal Self	2.52	2.11	2.21
		Self-Regard		
Mean	Real Self	10.62	11.14	10.71
SD	Real Self	2.80	2.48	2.65
Mean	Ideal Self	12.93**	12.94**	13.87**
SD	Ideal Self	1.89	1.92	1.71
		Self-Acceptance		
Mean	Real Self	15.05	15.37	14.44
SD	Real Self	4.00	3.28	3.39
Mean	Ideal Self	14.67	13.92*	13.60
SD	Ideal Self	3.36	3.01	2.64

TABLE 1 (*continued*)

Parameter	Instruction	Group I	Group II	Group III
		Nature of Man: Constructive		
Mean	Real Self	11.27	11.45	11.10
SD	Real Self	2.13	2.19	1.96
Mean	Ideal Self	12.15*	11.85	12.22**
SD	Ideal Self	2.22	1.94	1.76
		Synergy		
Mean	Real Self	6.28	6.10	6.25
SD	Real Self	1.47	1.68	1.39
Mean	Ideal Self	6.84*	6.83*	6.75**
SD	Ideal Self	1.34	1.33	1.34
		Acceptance of Aggression		
Mean	Real Self	16.33	14.78	15.41
SD	Real Self	2.93	3.54	3.46
Mean	Ideal Self	15.06*	14.19	14.41*
SD	Ideal Self	3.04	2.56	2.62
		Capacity for Intimate Contact		
Mean	Real Self	17.22	16.61	16.82
SD	Real Self	3.36	3.25	3.19
Mean	Ideal Self	17.25	15.42	16.41
SD	Ideal Self	3.63	3.32	3.81

* Significantly different from "Real Self" mean at .05 or .02 level.
** Significantly different from "Real Self" mean at .01 or .001 level.

D. Summary

Three groups of undergraduates completed a measure of self-actualization, the Personal Orientation Inventory, under both standard and ideal-self instructions. Contrary to earlier studies, which compared scores under standard *vs.* "good-impression" instructions, results were inconsistent across the 12 scales. Thus, on six scales subjects consistently appeared more self-actualized under ideal-self instructions than standard instructions, while on four others the reverse was true. Inner-Directedness (regarded by Knapp as the single most representative overall measure of self-actualization) and Capacity for Intimate Contact were not affected significantly by the different instructions in any of the groups.

References

1. Braun, J. R., & Asta, P. Intercorrelations between Personal Orientation Inventory and Gordon Personal Inventory scores. *Psychol. Rep.*, 1968, **23**, 1197-1198.
2. Braun, J. R., & LaFaro, D. A further study of the fakability of the Personal Orientation Inventory. *J. Clin. Psychol.*, 1969, in press.

3. DAMM, V. J. The relation of ego-strength to creativity and intelligence in high school students. Doctoral dissertation, University of Oregon, Eugene, Oregon. Ann Arbor, Michigan: Univ. Microfilms, 1967, No. 67-16, 154.
4. GROSSACK, M. M., ARMSTRONG, T., & LUSSIEV, G. Correlates of self-actualization. *J. Human. Psychol.,* 1966, **6**, 87.
5. ILARDI, R. L., & MAY, W. T. A reliability study of Shostrom's Personal Orientation Inventory. *J. Human. Psychol.,* 1968, **8**, 68-72.
6. KNAPP, R. R. Relationship of a measure of self-actualization to neuroticism and extraversion. *J. Consult. Psychol.,* 1965, **29**, 168-172.
7. RAPAPORT, G. M. "Ideal self" instructions, MMPI profile changes, and the prediction of clinical improvement. *J. Consult. Psychol.,* 1958, **22**, 459-463.
8. RUNYON, R. P., & HABER, A. Fundamentals of Behavioral Statistics. Reading, Mass.: Addison-Wesley, 1967.
9. SHOSTROM, E. L. A test for the measurement of self-actualization. *Educ. Psychol. Meas.,* 1965, **24**, 207-218.
10. ———. Manual, Personal Orientation Inventory (2nd ed.). San Diego, Calif.: Educ. & Indus. Testing Service, 1966.
11. SHOSTROM, E. L., & KNAPP, R. R. The relationship of a measure of self-actualization (POI) to a measure of pathology (MMPI) and to therapeutic growth. *Amer. J. Psychother.,* 1966, **20**, 193-202.
12. ZACCARIA, J. S., & WEIR, W. R. A comparison of alcoholics and selected samples of nonalcoholics in terms of a positive concept of mental health. *J. Soc. Psychol.,* 1967, **71**, 151-157.

Department of Psychology
University of Bridgeport
Bridgeport, Connecticut 06602

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 165-168.

THE CLINICAL UTILITY OF BLOCK'S EGO-RESILIENCY AND EGO-CONTROL SCALES*

Department of Psychology, Saint Louis University

DENNIS KNEFF, ROBERT T. BODENSTEINER, THOMAS W. VODDE,
AND MALCOLM D. GYNTHER

A. INTRODUCTION

Block (1) has conducted extensive research on the two major factors of the MMPI, which has taken into account both the need to control for the confounding influence of stylistic factors and the need to obtain behavioral correlates as validating evidence. Based on data from nonpsychiatric samples, he identified the first factor as "ego-resiliency" and the second as "ego-control." He also developed two scales to measure these dimensions in which the effects of acquiescence and social desirability (SD) are controlled.

Further information regarding the clinical utility of Block's ego-resiliency (ER-S) and ego-control (EC-5) scale scores seems desirable, especially since Block has indicated that his scales should operate with greater discriminating power "in samples containing individuals with a wider range of psychopathology—samples more appropriate to the initial intent and aspiration of the MMPI" (1, p. 109). The purpose of the present study was to investigate the relationships among ER-S and EC-5 scores of psychiatric inpatients and diagnostic classifications, age, sex, race, *IQ*, education, marital status, and other clinically relevant variables.

B. METHOD

1. *Pilot Study*

Norms for hospitalized psychiatric patients were obtained by selecting every eighth patient from an alphabetized list of 1,281 Malcolm Bliss Mental Health Center patients who had been given MMPI's between 1957 and 1968. As some data were lost due to the MMPI short forms, this procedure resulted in a sample of 130 patients: 21 Negro males, 53 white males, 11 Negro females, and 45 white females. The range of ER-S scores was nine to 31 with a mean of 18.56 and standard deviation of 5.59. The range of the

---

* Received in the Editorial Office, Provincetown, Massachusetts, on April 2, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

EC-5 scores was seven to 29 with a mean of 16.25 and standard deviation of 9.32. The respective distributions were both slightly positively skewed, but very close to normal.

A random sample of 40 *S*s was selected from the 130 *S*s to examine possible relationships between ER-S and EC-5 scale scores and number of previous psychiatric hospitalizations, and social maladjustment status (i.e., prisoner or court referral), in addition to the variables referred to earlier. No significant relationships were found, but several trends were noted: (*a*) behavior disorders tended to obtain relatively high ER-S scores and relatively low EC-5 scores; (*b*) neurotics and organics tended to obtain the opposite pattern of scores; and (*c*) no relationship between either score and psychotic diagnosis was observed.

2. *Main Study*

a. Subjects. *S*s were drawn from the entire pool of 1,281 patients by including every tenth name listed until 100 were obtained. *S*s used in the pilot study proper ($N = 40$) were excluded. The age range for this sample was 17 to 61 with a mean of 32.77 years. The *IQ*s ranged from 63 to 128 with a mean of 96.27, and the range of education was five to 20 years with a mean of 10.08.

b. Procedure. Each *S*'s MMPI was scored for the ER-S and EC-5 scales. The ER-S scale consists of 40 items, 20 keyed true, 20 keyed false, with an average SD rating of 5.16 on a nine-point scale. The EC-5 scale consists of 32 items, 16 keyed true, 16 keyed false, with an average SD rating of 5.33. For further details, see Block (1). *S*s were grouped into high *vs.* low scorers on each scale with the use of the norms developed in the earlier procedure. For ER-S, highs were defined as 18 and above, lows as 17 and below; for EC-5, highs were defined as 17 and above, lows as 16 and below. High and low ER-S and EC-5 scorers were then compared on the demographic, diagnostic, and clinical variables.

C. RESULTS AND DISCUSSION

Aside from diagnosis, which will be discussed in the following paragraph, the only patient variables that were significantly related to the categories derived from scores on Block's scales were age and sex differences for the ego-control dimension. Older *S*s obtained significantly higher scores than younger *S*s ($t = 2.01$, $p < .05$), which is consistent with Block's interpretation of low scores on the ego-control dimension as indicating impulsivity, a characteristic likely to decrease with age. The finding that females were significantly more controlled than males ($\chi^2 = 15.77$, $p < .001$) supports Block's similar finding which he has interpreted as reflecting culturally determined sex-role patterns concerning management of impulses. Block suggests that separate scales be developed for males and females.

The diagnostic classifications of high *vs.* low scorers differed significantly at less than the .05 level on both scales. With regard to the ego-resiliency dimension, the significant differences resulted from behavior disorders having significantly higher scores than neurotics and organics. For the ego-control dimension, on the other hand, behavior disorders had significantly lower scores than those found among neurotics and organics. Scale scores were unrelated to the psychotic diagnosis, a finding that is consistent with the data obtained in the pilot study.

These findings suggested that it might be instructive to run a secondary analysis with the use of a quadrant approach. To sharpen the differences, the 27 *S*s whose scores were at or near the midpoint of either scale were excluded. ER-S scores of 17 or less were classified as low, with ER-S scores of 20 and above classified as high. EC-5 scores of 15 or less were classified as low, with EC-5 scores of 18 and above classified as high. A comparison between these four groups on admission status approached significance ($\chi^2 = 7.53$; $.05 < p < .10$), primarily due to the fact that more first admission patients fell in the high ER-S—low EC-5 cell than would be expected by chance.

Because of an insufficient sample size, all organic and neurotic *S*s were eliminated from the analysis of diagnosis by quadrants. Results of this analysis approached significance ($\chi^2 = 7.16$; $.05 < p < .10$) reflecting the fact that there were more behavior disorders and fewer psychotics than expected with high ER-S, low EC-5 scores, while there were more psychotics and fewer behavior disorders with high ER-S, high EC-5 scores. Although there were only six neurotic patients in the total study, four of these patients were found in the low ER-S, high EC-5 category.

These results are obviously not very impressive. One could not predict diagnosis with any useful degree of accuracy by means of the ER-S score, the EC-5 score, or a combination of the two. Yet some significant relationships have been demonstrated, which is an improvement over the completely negative findings of Gynther and Brilliant (2) with the use of Welsh's (4) anxiety-repression scales. One might conclude that Block's unconfounded scales have more potential diagnostic utility than confounded measures of the same MMPI factors.

One suggestion for further research along these lines is to use Block's ER-O scale (108 items, also measuring the first factor, balanced for true-false, but selected without regard for SD scale values) in place of the ER-S scale. Kneff (3) found that psychiatric patients obtained significantly lower scores than Block's normal *S*s on the former scale, whereas the scores obtained in this study with the latter scale do not differ from scores Block obtained from normals. If ego-resiliency means "integrated and yet open to experience, able to cope resourcefully" (1, p. 110), one should use the form of the scale which most effectively separates normals and patients on this dimension.

D. SUMMARY

The purpose of this study was to investigate the relationships among Block's ego-resiliency and ego-control scale scores and diagnostic classification, age, sex, race, *IQ*, education, marital status, prisoner-nonprisoner status, and admission status. Norms for psychiatric inpatients were developed in a pilot study. The main sample of 100 patients was divided into high and low scorers on each scale. The diagnostic classifications of high and low scorers differed significantly on both scales, with neurotics and organics scoring higher than behavior disorders on the ego-resiliency dimension and the opposite pattern resulting for the ego-control dimension. Scale scores were, however, unrelated to the psychotic diagnosis. The only significant findings related to the other variables were that older patients scored higher than younger patients, and that females scored higher than males on the ego-control scale. This approach appears to be promising, but changes in procedure are necessary to increase the predictive power of these scales.

REFERENCES

1. BLOCK, J. The Challenge of Response Sets. New York: Appleton-Century-Crofts, 1965.
2. GYNTHER, M. D., & BRILLIANT, P. J. The diagnostic utility of Welsh's A-R categories. *J. Project. Tech. Personal. Assess.*, 1968, **32**, 572-574.
3. KNEFF, D. Personality characteristics of brain-damaged patients. Unpublished Master's thesis, Saint Louis University, Saint Louis, Missouri, 1969.
4. WELSH, G. S. Factor dimensions in *A* and *R*. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic Readings on the MMPI in Psychology and Medicine*. Minneapolis, Minn.: Univ. Minnesota Press, 1956.

Department of Psychology
Saint Louis University
221 North Grand Boulevard
Saint Louis, Missouri 63103

Published as a separate and in *The Journal of Psychology,* 1969, **72**, 169-177.

FACTORIAL STUDY AND ANALYSIS OF FACTOR VARIANCE OF GENERAL ADAPTABILITY SELF-RATINGS FROM A COMBINED SAMPLE OF MALE AND FEMALE SUBJECTS*[1]

Stevens Institute of Technology

RIO SCIORTINO

A. INTRODUCTION

In a recent study Sciortino (7) factorized the scores of a combined sample of male and female undergraduate *S*s on a Motivational Adjective List (MAL) and two factors were extracted: striving and assertion. The MAL was constructed by Sciortino (6), together with two other separate adjective lists, the General Adaptability Adjective List (GAAL) and the Intellective Adjective List (IAL), to cover three aspects of the intellectually superior and creative individual: namely, motivation, general adaptability, and intellection. When an analysis of factor variance (8) was performed on the two factors of the MAL the results showed that the female *S*s rated themselves significantly higher than the male *S*s on both factors: striving and assertion.

The purpose of this study was (*a*) first, to find the factorial structure of the General Adaptability Adjective List (GAAL) for a combined sample of male and female *S*s; and (*b*) second, to determine, by analysis of factor variance, the effect of sex, male *vs.* female *S*s, on the factor score means of the factors of the GAAL obtained in section (*a*). For the factorial structure of the GAAL several factorial methods were used and the one that, most adequately, met the criteria of simple structure was chosen for the interpretation of factors. No correlations between the scores on the GAAL and scores on other instruments were undertaken in this study. In the analysis of factor variance of the GAAL factors, F ratios were computed for the factor score means obtained from several factorial methods.

B. FACTOR ANALYSES

1. *Method*

a. Sample. Male ($N = 100$) and female ($N = 102$) undergraduate stu-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on April 7, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Data for this study were collected in the spring quarter of 1962 at the Ohio State University.

dents required to serve as *S*s to fulfill the experimental requirements for a course in general psychology were used.

b. Description of the test: The General Adaptability Adjective List (GAAL). The GAAL, constructed by Sciortino (6), consists of the 39 items listed in Table 1. *S*s rated themselves on each of the 39 items using the

TABLE 1
MATRIX OF MEANS, STANDARD DEVIATIONS, PROMAX-EQUAMAX FACTORS FOR SCORES ON THE GENERAL ADAPTABILITY ADJECTIVE LIST (GAAL) BY MALE ($N = 100$) AND FEMALE ($N = 102$) *S*s

			Promax-Equamax factors				
Variables	*M*	*SD*	A	B	C	D	E
1. Spontaneous	3.26	1.02	25	18	—25	07	19
2. Resilient	3.47	1.01	39	—09	10	23	—13
3. Expansive	3.20	.87	38	14	00	—18	07
4. Cordial	3.54	.82	35	—26	29	24	—01
5. Amorous	3.71	.97	49	05	—04	—09	03
6. Warm	3.67	.86	52	01	06	07	—11
7. Spirited	3.41	.84	08	42	13	20	—05
8. Lively	3.53	.84	03	46	06	35	—10
9. Enthusiastic	3.66	.72	08	31	26	05	02
10. Cheerful	3.58	.88	21	13	28	25	—12
11. Vigorous	3.45	.85	04	51	—08	05	09
12. Quick	3.32	.77	11	—08	—13	55	09
13. Energetic	3.44	.82	—06	63	—08	09	03
14. Brisk	2.95	.91	—07	19	—10	36	06
15. Active	3.72	.79	13	51	—05	—01	—09
16. Wholesome	3.66	.83	02	22	43	04	06
17. Sprightly	3.22	.90	—09	31	13	39	00
18. Healthy	4.03	.85	—18	25	31	20	04
19. Dynamic	3.15	.82	—03	52	07	08	06
20. Agile	3.37	.97	02	22	06	37	—05
21. Fluent	2.89	.88	01	—07	—12	13	59
22. Facile	3.09	.74	02	06	—02	35	13
23. Evocative	3.16	.87	18	—05	05	02	41
24. Depictive	2.56	.83	—16	01	00	03	50
25. Articulate	2.92	.83	—16	04	11	—03	59
26. Persuasive	3.43	.91	22	06	—02	07	20
27. Felicitous	3.24	.88	36	07	00	21	05
28. Eloquent	2.78	.81	01	12	00	—02	45
29. Communicative	3.39	.97	37	04	—07	12	14
30. Tactful	3.13	.97	—02	—32	16	18	32
31. Responsive	3.46	.88	36	08	16	—21	13
32. Leading	2.95	.93	—07	10	—04	27	26
33. Cooperative	3.67	.77	05	—01	53	03	—05
34. Attentive	3.46	.77	02	—13	56	—06	08
35. Tolerant	3.62	.97	22	—05	44	—13	—01
36. Sociable	3.65	.93	53	00	00	14	—09
37. Responsible	3.79	.83	—09	06	46	—03	00
38. Empathetic	3.26	.97	16	—04	21	—24	34
39. Congenial	3.72	.84	53	—05	18	10	—10
.10 h.c.			18	20	20	18	23

Note: Initials h.c. stand for hyperplane count. Decimal points omitted.

following five-point scale: (*a*) least, (*b*) less than moderately, (*c*) moderately, (*d*) more than moderately, (*e*) most. The self-ratings constituted the scores of *S*s on each of the 39 items. All the adjectives were defined and collected in a separate booklet entitled Adjective Definitions.

c. Procedure. The GAAL and the Adjective Definitions booklet were given to *S*s at the same time. *S*s were asked to consult the Adjective Definitions booklet when they were unsure about the meaning of certain adjectives. The same examiner administered the GAAL to both male and female groups on two different occasions. The findings reported here are based on the combined groups ($N = 202$).

2. *Results*

a. Computations.[2, 3] The computations for this study were carried out on the IBM 7094 operated by the Princeton Computer Center and with the P-Stat computer program.

The iterative factor analysis routine of the P-Stat computer program was used to perform a factor analysis on the matrix of item correlations. The routine cycles on the number of significant roots and then iterates on the stability of diagonal estimates. The obtained initial matrix of significant factors corresponds to the significant roots. Table A shows the matrix of item correlations.

Three orthogonal rotations of the initial matrix of significant factors were performed with the use of the quartimax (4), the varimax (2), and the equamax (5), procedures. Table B shows the matrix of the initial significant factors, and of the quartimax, varimax, and equamax factors. Three promax rotations (1), with $K = 4$, were performed on each of the three orthogonal rotations obtaining the promax-quartimax, the promax-varimax, and the promax-equamax solutions. Table C shows the matrix of the promax-quartimax and promax-varimax factors, and Table D shows their respective correlations.

Table 1 shows the means and standard deviations of the GAAL items, and the promax-equamax factors for the combined sample. Table 2 shows the correlations of the promax-equamax factors. Table B and C, of the NAPS doc-

---

[2] The following materials have been deposited with the National Auxiliary Publication Service: Table A, matrix of item correlations (39 × 39) of the GAAL; Table B, matrix of initial significant factors, and of quartimax, varimax, and equamax factors; Table C, matrix of promax-quartimax and promax-varimax factors; and Table D, matrix of correlations of promax-quartimax and promax-varimax factors. Order NAPS Document No. 00419 from National Auxiliary Publication Service, c/o CCM Information Sciences, Inc., 22 West 34th St., New York, New York, 10001; remitting in advance $3.00 for photocopies or $1.00 for microfiche.

[3] The data matrices of the male ($N_1 = 100$) and female ($N_2 = 102$) *S*s on the 39 items (M) of the GAAL were combined into an $N_1 + N_2$ by M table to obtain the $M \times M$ correlation matrix, the mean scores, and the standard deviations.

ument, and Table 1, of the present study, show that the total .10 hyperplane count of the promax solutions is higher than that of the orthogonal solutions. However, since the promax-equamax solution has the highest total .10 hyperplane count, it was used to interpret the factors.

TABLE 2
CORRELATION MATRIX OF PROMAX-EQUAMAX FACTORS FOR SCORES ON THE GENERAL ADAPTABILITY ADJECTIVE LIST (GAAL)
(Combined Samples ($N = 202$))

	Promax-Equamax factors			
Factors	A	B	C	D
Sociability (A)				
Energy (B)	42*			
Attentiveness (C)	46*	36*		
Quickness (D)	30*	40*	33*	
Articulateness (E)	48*	38*	26*	26*

Note: Decimal points omitted.
* $p = .01$.

b. Interpretation of factors. Interpretation of factors rests on loadings of .25 or greater ($p = .01$ if $N = 202$). A factor whenever possible was assigned the name of the variable that has the highest loading with it.

Factor A (Sociability) has loadings arranged as follows: sociable, .53; congenial, .53; warm, .52; amorous, .49; resilient, .39; expansive, .38; communicative, .37; responsive, .36; felicitous, .36; cordial, .35; and spontaneous, .25. This factor refers to *S*s' sociability or disposition to enjoy the company of others.

Factor B (Energy) has loadings arranged as follows: energetic, .63; dynamic, .52; vigorous, .51; active, .51; lively, .46; spirited, .42; tactful, —.32; enthusiastic, .31; sprightly, .31; cordial, —.26; and healthy, .25. This factor refers to *S*s' energy level. However, the negative loadings of items tactful and cordial suggest that *S*s' manifestation of energy detracts from, or has a negative relationship with, their being cordial and tactful toward others.

Factor C (Attentiveness) has loadings arranged as follows: attentive, .56; cooperative, .53; responsible, .46; tolerant, .44; wholesome, .43; healthy, .31; cordial, .29; cheerful, .28; enthusiastic, .26; and spontaneous, —.25. This factor refers to *S*s' attentiveness or consideration for others. However, the negative loading of item spontaneous indicates that *S*s' attentiveness is rather controlled and lacks, therefore, a certain degree of spontaneity.

Factor D (Quickness) has loadings arranged as follows: quick, .55; sprightly, .39; agile, .37; brisk, .36; lively, .35; facile, .35; leading, .27; and cheerful, .25. This factor refers to *S*s' quickness, or agility in movement.

Factor E (Articulateness) has loadings arranged as follows: articulate, .59; fluent, .59; depictive, .50; eloquent, .45; evocative, .41; empathetic, .34; tactful, .32; and leading, .26. This factor refers to *S*s' articulateness, or skill to express themselves clearly and easily.

C. Analysis of Factor Variance

1. *Method*

Computations:[4] The orthogonal and oblique factors of the GAAL, reported above, were used in this section to obtain their respective factor scores. For this purpose the factor score subroutine of the P-Stat computer program, based on the multiple regression method, was used. The obtained factor scores were then subjected to a one-way classification analysis of variance by using another subroutine of the P-Stat computer program. Table 3 shows the factor score means, standard deviations, and *F* ratios of the male and female *S*s on the five factors of the GAAL (sociability, energy, attentiveness, quickness, and articulateness), and for six factorial procedures (quartimax, varimax, equamax, promax-quartimax, promax-varimax, and promax-equamax).

2. *Results*

a. Factor A (Sociability). In all six factorial procedures the means of the female *S*s are higher than those of the male *S*s. However, the *F* ratios reached statistical significance only in the equamax and promax-equamax solutions, beyond the .01 level, and in the promax-varimax solution, at the .05 level. The significant *F* ratios indicate that the female *S*s rated themselves as being more sociable—more disposed to enjoy the company of others—than the male *S*s.

b. Factor B (Energy). In all six factorial procedures, the means of the female *S*s are higher than those of the male *S*s. However, the *F* ratios reached statistical significance only in the quartimax, varimax, promax-quartimax, and promax-varimax solutions, at the .01 level, and in the promax-equamax solution, at the .05 level. The significant *F* ratios indicate that the female *S*s rated themselves as having a higher energy level than the male *S*s.

c. Factor C (Attentiveness). In the varimax, equamax, promax-quartimax, promax-varimax, and promax-equamax solutions the means of the female *S*s are higher than those of the male *S*s; however, in the quartimax solution the mean of the male *S*s is higher than that of the female *S*s. The *F*

---

[4] $S = Z R^{-1} F$; where S = factor score matrix, Z = matrix of observed scores in standard score form, R = intercorrelation matrix, and F = rotated factor loading matrix.

TABLE 3

MATRIX OF MEANS, STANDARD DEVIATIONS, AND *F* RATIOS FOR SCORES ON THE FIVE FACTORS OF THE GENERAL ADAPTABILITY ADJECTIVE LIST (GAAL) AND FOR SIX FACTORIAL PROCEDURES—QUARTIMAX, VARIMAX, EQUAMAX, PROMAX-QUARTIMAX, PROMAX-VARIMAX, AND PROMAX-EQUAMAX (Males ($N = 100$) Females ($N = 102$))

Factorial procedures	A Sociability		B Energy		C Attentiveness		D Quickness		E Articulateness	
	M	F	M	F	M	F	M	F	M	F
Quartimax										
M	4.99	5.01	4.48	5.51	5.05	4.95	4.81	5.18	5.12	4.88
SD	1.18	1.06	.93	1.02	1.20	1.13	1.16	1.19	1.21	1.30
F	.01		55.51**		.41		5.02*		1.85	
Varimax										
M	4.97	5.03	4.51	5.48	4.74	5.25	5.01	4.99	5.11	4.89
SD	1.08	1.12	1.00	1.04	1.09	1.20	1.17	1.14	1.25	1.23
F	.18		45.52**		10.15**		.01		1.63	
Equamax										
M	4.52	5.46	4.93	5.07	4.71	5.28	5.09	4.91	4.98	5.01
SD	1.04	1.05	1.09	1.18	1.06	1.19	1.24	1.16	1.17	1.14
F	41.09**		.71		12.85**		1.12		.03	
Promax-Quartimax										
M	4.82	5.18	4.33	5.65	4.83	5.17	4.67	5.32	5.12	4.88
SD	1.29	1.31	1.15	1.33	1.28	1.38	1.12	1.27	1.27	1.33
F	3.84		56.66**		3.20		15.08**		1.69	
Promax-Varimax										
M	4.80	5.20	4.35	5.64	4.60	5.39	4.83	5.16	5.03	4.96
SD	1.25	1.36	1.14	1.30	1.14	1.34	1.26	1.34	1.37	1.30
F	4.85*		56.18**		19.95**		3.30		.14	
Promax-Equamax										
M	4.36	5.63	4.79	5.21	4.58	5.41	5.00	5.00	4.83	5.17
SD	1.15	1.28	1.20	1.36	1.13	1.34	1.35	1.25	1.27	1.33
F	54.78**		5.49*		22.69**		.00		3.38	

* $p = .05$.
** $p = .01$.

ratios reached statistical significance only in the varimax, equamax, promax-varimax, and promax-equamax solutions, and beyond the .01 level. The significant F ratios indicate that the female Ss rated themselves as being more attentive, or more considerate toward others, than the male Ss.

d. Factor D (Quickness). In the quartimax, promax-quartimax, and promax-varimax solutions the means of the female Ss are higher than those of the male Ss. In the varimax and equamax solutions the means of the male Ss are higher than those of the female Ss, and in the promax-equamax solution the mean of the male Ss is equal to that of the female Ss. The F ratios reached statistical significance only in the quartimax solution, at the .05 level, and in the promax-quartimax solution, at the .01 level. The significant F ratios indicate that the female Ss rated themselves as being quicker, or more agile in movement, than the male Ss.

e. Factor E (Articulateness). The mean of the male Ss is higher than that of the female Ss in the quartimax, varimax, promax-quartimax, and promax-varimax solutions, but is lower than the mean of the female Ss in the equamax and promax-equamax solutions. Since no solution yielded a statistically significant F ratio, no sex difference was found for factor articulateness.

D. Discussion and Conclusion

The correlations of the promax-equamax factors shown in Table 2 range from .26 to .48. They are all statistically significant, at the .01 level, and indicate that the five factors of the GAAL, sociability, energy, attentiveness, quickness, and articulateness share a certain degree of variance. Factor sociability refers to Ss' disposition to enjoy the company of others. Factor energy refers to Ss' energy level. Factor attentiveness refers to Ss' consideration for others. Factor quickness refers to Ss' agility in movement, and factor articulateness refers to Ss' skill to express themselves clearly and easily.

The F ratios of the five factors of the GAAL, shown in Table 3, indicate that the female Ss rated themselves significantly higher than the male Ss on factors sociability, energy, attentiveness, and quickness. No significant sex difference was found for factor articulateness.

In performing an analysis of factor variance, Pinneau *et al.* (3) have recommended the use of uncorrelated factor scores; however, in this study both uncorrelated and correlated factor scores were used. Since the orthogonal solutions, as a group, and their corresponding oblique solutions, as another group, yielded significant F ratios on the same four of the five factors of the GAAL, either group of solutions could have been used to find out the sex differences reported here. However, only the promax-varimax and the promax-equamax, both oblique solutions, yielded significant F ratios on three different factors: sociability, energy, and attentiveness. Moreover, it was only the quartimax and promax-quartimax solutions that yielded significant F ratios on factor quickness.

The results of Table 3 do not show, unambiguously, which of the factorial procedures, either orthogonal or oblique, was most effective in detecting sex differences. Moreover, whether the statistical significances reported for four of the five factors of the GAAL—sociability, energy, attentiveness, and quickness—could be attributed to such sources as *S*s' social expectations, or their personal preferences, or to the factorial procedures used here cannot be ascertained from the results of this exploratory study. Other studies must be carried out, using different samples of male and female *S*s, to investigate (*a*) the factorial invariance of the GAAL; (*b*) the factorial structure of the GAAL under different test instructions; and (*c*) the influence of different orthogonal and oblique rotations on the estimation of factor scores, and on the statistical significance of the difference between pairs of factor score means.

E. Summary

An iterative factor analysis was performed on the scores obtained from a General Adaptability Adjective List (GAAL) for a combined sample of undergraduate male and female *S*s ($N = 202$). The initial matrix of significant factors was rotated according to the quartimax, varimax, and equamax procedures. The promax rotation was, then, performed on each of the three orthogonal solutions obtaining the promax-quartimax, promax-varimax, and promax-equamax solutions. Interpretation of factors was based on the promax-equamax solution and the factors obtained were sociability, energy, attentiveness, quickness, and articulateness.

A simple analysis of factor variance of the factor score means of male and female *S*s, on the five factors of GAAL mentioned above, yielded the following results. The female *S*s rated themselves significantly higher than the male *S*s on factors sociability, energy, attentiveness, and quickness; no significant sex difference was found for factor articulateness. Factor scores were estimated by the multiple regression method and from the quartimax, varimax, equamax, promax-quartimax, promax-varimax, and promax-equamax procedures.

References

1. HENDRICKSON, A. E., & WHITE, P. O. Promax: a quick method for rotation to oblique simple structure. *Brit. J. Stat. Psychol.*, 1964, **17**, 65-70.

2. KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika,* 1958, **23**, 187-200.
3. PINNEAU, S. R., LEVINE, A. J., SCHURR, B. C., & BUTLER, D. C. Analysis of factor variance: One-way classification. *Percept. & Motor Skills,* 1966, **23**, 1209-1210.
4. SAUNDERS, D. R. An analytic method for rotation to orthogonal simple structure. Research Bulletin, 53-10, Educational Testing Service, Princeton, New Jersey, 1953.
5. ———. Trans-varimax: Some properties of the ratiomax and equamax criteria for blind orthogonal rotation. Paper read at the American Psychological Association Convention, St. Louis, Missouri, 1962.
6. SCIORTINO, R. Personality and Creative Abilities. Doctoral Dissertation, Ohio State University, Columbus, Ohio, 1963. No. 64-6958.
7. ———. Factorial analysis of motivational self-ratings from a combined sample of male and female subjects. *Psychol. Rep.,* 1967, **21**, 573-576.
8. ———. Analysis of factor variance of motivational self-ratings by male and female subjects. *J. of Psychol.,* 1968, **69**, 169-174.

Stevens Institute of Technology
Hoboken, New Jersey 07030

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 179-181.

THE FEAR OF DEATH AND THE FEAR OF DYING*[1]

*Department of Psychology, Wellesley College and
Suicide-Prevention and Crisis Service, Buffalo*

Lora-Jean Collett and David Lester

A. Introduction

Investigators of attitudes toward death have realized the need for standardized measures of the fear of death (2). Sarnoff and Corwin (4), Boyar (1), and Lester (3) have devised measures of the fear of death and investigated the reliability and validity of their measures. However, the items in their scales cover a wide variety of themes: fear of dying, fear of being dead, aversion to funerals, and even worry over the death toll on highways.

The purpose of the present study was to devise separate measures of death fears, attempting to distinguish between the fear of death from the fear of the process of dying and to differentiate between these fears depending upon whether they are for oneself or for another.

B. Method

The *S*s formed an original sample (A) and a replication sample (B) each consisting of 25 female undergraduates.

Thirty-eight statements were composed concerning the four fears: fear of death of self, fear of death of others, fear of dying of self, and fear of dying of others. The number of items in each subscale was 10, 10, 8, and 10 respectively.[2]

The items were given to the *S*s who were required to indicate their agreement or disagreement with each item on a 6-point scale ranging from strong agreement (+3) to strong disagreement (—3).

C. Results

The scores of the *S*s in sample A on each item were correlated with the

---

* Received in the Editorial Office, Provincetown, Massachusetts, on April 9, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Address reprint requests to David Lester, Suicide-Prevention and Crisis Service, 312 Lafayette Building, Buffalo, New York 14203.

[2] A copy of the scale may be obtained from The National Auxiliary Publications Service. Order NAPS Document No. 00418 from ASIS National Auxiliary Publications Service, c/o CCM Information Sciences, Inc., 22 West 34th Street, New York, New York 10001; remitting in advance $3.00 for photocopies or $1.00 for microfiche.

total score for the subscale to which the item belonged. All items whose correlations were not significant at the .10 level of significance (i.e., the correlation was less than .26) were eliminated. Six items were eliminated as a result.[3]

The intercorrelations between the four subscales were calculated and these ranged from .03 to .58 and were, in general, low especially where the type of fear and the referent of the fear differed. For example, fear of dying of self and fear of death of others had a zero correlation (.03 and —.07 in the two samples), whereas fear of dying of others and fear of death of others had a significant but low correlation (.46 and .40 in the two samples, two-tailed $p < .05$).[4]

A three-way analysis of variance for repeated measures was carried out on the data (5) with the two samples treated as a replication factor.[5] The terms of the analysis involving the replication factor were all nonsignificant. The largest F-ratio was 1.16 ($df = 1{,}48$).

The *S*s showed a significantly higher fear of death than of dying ($F = 46.53$; $df = 1{,}48$; $p < .01$). They also showed a significantly greater fear when the self was the referent than when another was the referent ($F = 14.74$; $df = 1{,}48$; $p < .01$). The interaction of death/dying with self/other was also significant ($F = 23.78$, $df = 1{,}48$; $p < .01$). The mean scores per item for each of the subscales (a measure used to control for the length of each scale) were fear of death of self, .48; fear of death of others, .59; fear of dying of self, .33; and fear of dying of others —.78.

D. DISCUSSION

The low intercorrelations between the four fears (fear of death of self, fear of dying of self, fear of death of others, and fear of dying of others) indicate the potential usefulness of differentiating these four specific fears rather than indiscriminately grouping all items in the same scale. Such a step would lead to an improvement in available measures of the fear of death. The low correlations may reflect a general death fear which is tapped by all death scales, and in cases where the scales share a common process (death/dying) or a common referent (self/other) they may reflect this shared object.

---

[3] The test used was a one-tailed test of significance.

[4] The intercorrelations for the two samples were death self/death others .22 and .26, death self/dying self .24 and .41, death self/dying others .09 and .58, death others/dying self .03 and —.07, death others/dying others .46 and .40, and dying self/dying others .28 and .40. A correlation greater than .39 is necessary for significance at the .05 level (two-tailed test).

[5] The interaction terms in analysis of variance were homogeneous (5).

Of the four fears, fear of dying of others was least feared. In fact for this subscale the mean score was negative, indicating that *S*s faced this topic in general rather than avoiding it. The differences in scores on the four scales are not a result of the different lengths of the four scales, since this was controlled for in the analysis of variance by using the mean score per item for each scale and each *S*. However, the differences could be a result of differences in the strength of the wording of the items of different scales. Such an occurrence would not be of importance when other variables are being compared to scores on the four scales.

E. SUMMARY

Scales to assess fear of death of self, fear of dying of self, fear of death of others, and fear of dying of others were devised. The low correlations between scores on these scales indicate the potential usefulness of differentiating between these four fears.

REFERENCES

1. BOYAR, J. I. The construction and partial validation of a scale for the measurement of the fear of death. *Diss. Abst.*, 1964, **25**, 2041.
2. LESTER, D. Experimental and correlational studies of the fear of death. *Psychol. Bull.*, 1967, **67**, 27-36.
3. ———. Fear of death of suicidal persons. *Psychol. Rep.*, 1967, **20**, 1077-1078.
4. SARNOFF, I., & CORWIN, S. M. Castration anxiety and the fear of death. *J. Personal.*, 1959, **27**, 274-285.
5. WINER, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

Suicide-Prevention and Crisis Service
312 Lafayette Building
Buffalo, New York 14203

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 183-188.

SOCIAL ROLE PERCEPTIONS OF CHILDREN IN HUTTERITE COMMUNAL SOCIETY*[1]

Department of Psychology, University of Manitoba

SHIRIN SCHLUDERMANN AND EDUARD SCHLUDERMANN

A. INTRODUCTION

The Hutterite society is four hundred years old. The three main tenets of the society are (*a*) Anabaptist ideology; (*b*) collectivist ideology; and (*c*) the principle of realization: i.e., commitment to translating ideology into action. Like other communal societies—e.g., the Kibbutzim—the Hutterite societies are agricultural communities. Hutterites firmly believe in the principle and practice of sharing everything equally according to needs. In the Hutterite societies cooperation, equality, fraternity, justice, freedom from competition and exploitation are some of the basic principles espoused. In keeping with their ideology, the Hutterite men and women wear distinctive uniform dress, dwell in colonies consisting of approximately 100 members, live in similar type dwelling houses built by themselves, eat in a common dining room the same food as everyone else, and abolish all visible signs of status symbol usually associated with rank, role, and position in society. Interested readers may read books by Victor Peters (6), John Horsch (3) and Paul Gross (2) to gain interesting and relevant information on social structure and organization, economic life, and religious beliefs of the Hutterites. At present, Hutterite colonies are located in Alberta, Manitoba, Saskatchewan, South Dakota, and Montana.

The child rearing practices in the communal societies involve "collective education." The responsibility for care, training, and discipline of children becomes the moral, economic, and ideological responsibility of the community. The specially trained members of the society assume much of the parental responsibility for educating the children. Collective upbringing in the Hutterite society has the following features:

During the first six weeks of his life, the baby is looked after by a female

---

* Received in the Editorial Office, Provincetown, Massachusetts, on April 9, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This research was supported by a grant from the Canada Council, and also assisted by a grant from the Faculty of Graduate Studies and Research, University of Manitoba. The statistical analyses were done at the University of Manitoba Computer Center.

member of the community. Thereafter the mother takes full charge of the baby. The first three years of the child's life are spent in the special care of his mother supplemented by abundant care by other close relatives and community members. Children attend the kindergarten between the ages of three to six years. A special kindergarten woman (mostly a grandmother) is in charge of these children for most of the day. Between the ages of six to 16 years, children attend the public school which is on the colony. The school day is supplemented by one hour of instruction in German and the Bible in the morning and evening. The German school teacher supervises and disciplines the school-age children. The colony leaders—i.e., the minister, the business manager, the German teacher, the kitchen woman, etc.—are also jointly responsible for the proper conduct and discipline of the child. After the age of 16 boys and girls leave school and take their sex-appropriate adult roles in the Hutterite colony. Since the colonies are of manageable size, face to face interaction and constant supervision of colony members are possible.

It was one of the aims of this investigation to examine the suggested perceptual regularities and universalities in interperson perception as mentioned by Dubin and Dubin (1) and others. It will be of interest to examine if these perceptual constancies in varied cultural contexts change qualitatively or quantitatively; and if only the latter, what might be the extent of such quantitative changes as estimated by psychological instruments (i.e., Semantic Differential). How basic and universal are the trends observed in social role differentiation and perception in a contrasting cultural milieu? The second objective of the study was to determine the nature and number of the basic dimensions underlying the perception of people by eight to 13-year-old boys and girls. How do the variables of age and sex influence the children's ability to perceive different social roles?

B. Method

1. *Subjects*

Three hundred and fifty-nine subjects were selected from 20 Hutterite colonies in Manitoba. The *S*s were eight- to 13-year-old boys and girls currently enrolled in the public schools in Hutterite colonies. The *S*s were selected on the basis of their school achievement records and teachers' evaluation. These children ranged in their abilities from average to somewhat above-average. In order to control somewhat for the effects of extreme differences in intelligence on conceptualization, very dull and outstanding students were eliminated from the sample. Socioeconomically, the Hutterite colonies form a homogenous social group located in rural areas.

2. *The Instrument, Concepts, and Scales*

Osgood and Osgood *et al.* (4, 5) have reported a series of cross-cultural studies with a variety of semantic scales, and have also rated a great variety of concepts. The factor structure of the semantic space has consistently remained stable across the varied language-culture groups studied. It appeared then that a well-standardized and tested tool in a great variety of cross-cultural and cross-linguistic groups would be an ideal tool for cross-cultural studies.

The social role concepts that were rated on a set of 15 bipolar semantic scales were as follows:

1. Family roles (males): Grandfathers, fathers, brothers.
2. Family roles (females): Grandmothers, mothers, sisters.
3. Occupations (males): Farm foremen, pigmen, doctors.
4. Occupations (females): Kitchen women, garden women, kindergarten teachers.
5. Self-concepts: Me (as I am), Me (as I would like to be).
6. Friends: Boy-friends, girl-friends.

The set of 15 bipolar semantic scales consisted of weak-strong, important-common, slow-fast, interesting-boring, afraid-brave, busy-lazy, dull-bright, useful-useless, sad-pleasant, soft-hard, unfriendly-friendly, gentle-rough, dirty-clean, active-quiet, small-large. The selection of scales was based on relevance, simplicity, representational character, and empirical variance. The scales were subjected to a series of successive empirical tests before the final selection was made. Care was taken to alternate the polarity direction of the semantic scales. Children rated the concepts on a five-point scale.

3. *Investigative Procedure*

The semantic differential was administered in small groups for the different age-groups separately. The procedure of administration, the forms of instruction, and the presentation of materials (Form II) were identical for all age groups and were kept exactly the same as those described by Osgood *et al.* (5).

C. RESULTS

The male and the female subjects were subdivided into two age subgroups (*a*) eight years to ten years, (*b*) 11 years to 13 years. Correlation matrices of 15 × 15 scales for the four age-sex subgroups were calculated separately, and factor analyzed by the principal axes method. The axes were rotated orthogonally according to the Varimax criterion.

Table 1 indicates that only the first factor in each subgroup accounts for a sufficiently large percentage of the variance to be considered as a common factor. All other factors in the groups are only specific factors. The principal factor matrix showed that 13 out of 15 scales had substantial loadings (i.e., greater than .40) on the first factors. There are only two or three scales with any substantial loadings on the second factors and only one or two scales on the third factors. In all subgroups, even though the first factor can be called a common factor, it does not account for more than one-third of the total variance; therefore, the factor space cannot be interpreted to be a unidimensional space. The failure to obtain one or two or more common factors casts doubt on the initial assumption that a few basic dimensions underlie eight to 13-year-old children's perception of social roles in the Hutterite colonies. In view of the results obtained in previous studies (5, 4, 7), one may question as to why no stable common factor space has emerged. The finding is all the more challenging in view of the fact that the investigators in their previous study (7), have used the same semantic scales and the same social roles as were used in this study, and have obtained a stable semantic space (i.e., Evaluative, Potency, and Activity) as reported by Osgood (5, 4).

TABLE 1
PROPORTION OF VARIANCE ACCOUNTED FOR BEFORE ROTATION

Subgroups	Ages	F_1	F_2	F_3	F_4	Cumulative percent variance
Young girls	8-10 yrs.	28.33	10.02	7.07	6.08	51.99
Young boys	8-10 yrs.	29.80	10.07	7.50	6.00	53.38
Older girls	11-13 yrs.	33.04	10.54	6.92	6.22	56.73
Older boys	11-13 yrs.	33.40	11.61	6.95	5.90	57.86

D. DISCUSSION

1. *Factor Structure*

As is seen from Table 1, the first three factors accounted for less than 50 percent of the total variance in each subgroup. This low common factor variance reflects generally low intercorrelations between the 15 semantic scales and could be accounted for in different ways: Firstly, there is the possibility that the subjects failed to understand the instructions clearly. Secondly, the *S*s may have had a somewhat vague understanding of the descriptive adjectives and they may have been uncertain in the use of these adjectives in the description of people. During the administration of the SD, special steps were taken to eliminate, as far as possible, these two sources of error. Peculiarities of the social environment in which Hutterite children grow up provide a very important third source of influence which would lower the common factor variance. Hutterite children of the age range studied usually have intimate contact with only one individual who fills a given social role in their own colony (e.g., farm foreman). It is therefore quite possible that the personal characteristics of the individual who holds a given office on their own colony influences the children's ratings of general social roles.

In the principal factor matrix of all the four subgroups, the first factor appeared as a global undifferentiated perception of the various social roles. Very likely the young respondents looked at the bipolar adjectival semantic scales in an essentially good-bad dimension. It is good to be brave, strong, interesting, useful, clean, etc. If persons possess these qualities they are "good," if not, then they are "bad." The next two factors (which can be regarded as specific factors), show interesting trends. The semantic scales "strong," "hard," and "rough" which compose F_2 in all four subgroups, appear repeatedly as the Potency (P) dimension in the cross-cultural researches of Osgood (4). The results obtained from the Hutterite children show some observable trends suggesting the early stages in the development toward a clear-cut E (Evaluative) P (Potency) A (Activity) factor structure. It is likely that, as the young respondents of this study grow older, they may be able to make the kind of differentiation in their perceptions that researchers using the semantic differential are accustomed to obtaining.

2. *Social Perception*

It may be said that the greater the number of common factors used by a group of respondents to describe social roles, the greater will be the cognitive differentiation involved. When the results of this study are compared with the results of studies (using the same social roles and the same semantic scales) on American children of the same age range (e.g., 7), the social perception of Hutterite children appears to be quite undifferentiated. In this section an attempt will be made to suggest some characteristics of Hutterite society which would delay the development of differentiated social perception among Hutterite children.

Hutterite children have considerable difficulty in noticing differences in social roles, because in Hutterite society all visible status symbols (which are the most obvious cues available to American children in the perception of social differences) are deliberately abolished. Hutterite society does not reward high achievement by way of distinguishable high status position. Hutterite ideology deliberately discourages the tendency to look at certain social roles, and certain types of work as superior or inferior. Since attaching greater value to certain social roles over others does not receive any overt positive reinforcement, Hutterite children have few incentives to pay special attention to characteristic features of different social roles. It is suggested that the development of differentiated social perception among Hutterite children is delayed because Hutterite children have fewer observable cues for role differentiation and because they have less encouragement for learning the differences between social roles.

E. Summary

Three hundred and fifty-nine *S*s, eight to 13 years old, selected from 20 Hutterite colonies in Manitoba, were asked to describe 16 social roles on 15 semantic differential scales. Factor analyses of the 15 × 15 four subgroups' correlation matrices yielded only one common factor which accounted for only one-third of the total variance; the rest of the factors emerged as specific variances. No clear-cut indication of a stable EPA semantic space was obtained for any of the age-sex groups. The assumption of a few basic dimensions underlying Hutterite children's perception of social roles became doubtful.

References

1. Dubin, R., & Dubin, E. R. Children's social perception: a review of research. *Child Devel.*, 1965, **36**, 809-838.
2. Gross, P. S. The Hutterian Way. Saskatoon, Canada: Freeman, 1965.
3. Horsch, J. The Hutterian Brethren—1528-1931. Goshen, Ind.: Mennonite Historical Society, 1931.
4. Osgood, C. E. Studies in the generality of affective meaning systems. *Amer. Psychol.*, 1962, **17**, 10-28.
5. Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. The Measurement of Meaning. Urbana, Ill.: Univ. Illinois Press, 1958.
6. Peters, V. All Things Common. The Hutterian Way of Life. Minneapolis, Minn.: Univ. Minnesota Press, 1965.
7. Schludermann, S., & Schludermann, E. A comparative factorial analysis of semantic structures in monolingual Indian and American school children. *Can. J. Psychol.*, 1969, **23**(2), 79-93.

Department of Psychology
The University of Manitoba
Winnipeg, Manitoba

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 189-194.

PEER INFLUENCES ON PRESCHOOL CHILDREN'S WILLINGNESS TO TRY DIFFICULT TASKS*[1]

Department of Educational Psychology, University of Georgia

E. PAUL TORRANCE

A. INTRODUCTION

Some children apparently fail to learn because they are unwilling to try more and more difficult tasks. Others fail because they compulsively insist upon trying tasks too difficult for them to master. Thus, to guide a child's learning effectively a teacher must be able to influence his willingness to try tasks of appropriate difficulty.

In another source, the author (4, 5) has presented evidence to indicate that learners differing in tendencies to attempt the difficult need different kinds of feedback concerning their performance. Learners low in willingness to attempt the difficult appear to need evaluative feedback and external criteria. Those high in willingness to attempt the difficult seem to be hampered by this kind of feedback and need opportunities to test their own limits and push their limits forward. Thus, teachers should be aware of individual differences in willingness to attempt the difficult and to differentiate feedback experiences accordingly.

Almost all personality studies of productive, creative people show that such people enjoy a calculated risk in which success or failure depends upon their own ability (6). In fact, there seems to be a positive relationship between willingness to attempt difficult tasks and achievement. Studies of the lives of eminent people who have made important social contributions suggest that one of the essentials for high achievement is a strong tendency to attempt difficult tasks. Instead of just adapting or adjusting to their environment, they deliberately exposed themselves to possible failure and committed themselves to goals that required sustained expenditures of intellectual and physical energy and continued changes in behavior. Thus, it would appear

---

* Received in the Editorial Office, Provincetown, Massachusetts, on April 9, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The research reported herein was performed as part of the activities of the Research and Development Center in Educational Stimulation, University of Georgia, pursuant to a contract with the United States Department of Health, Education, and Welfare, Office of Education, under provisions of the Cooperative Research Program.

that it is important to society that schools develop in children the ability to estimate their limits and a willingness to attempt difficult tasks.

The present study was conducted as a part of a sequence of activities to teach children to estimate more accurately their limits. It was designed specifically to determine the influence of three different peer conditions on willingness to attempt difficult tasks. The design of the study was suggested by the author's observations in working with United States Air Force personnel in advanced survival training that men were more likely to attempt difficult tasks and to test their limits in pairs than in individual or large group situations. The design of this study was also influenced by the finding of Miller, Bensen, Seidman, and Meeland (2) that men have a greater tolerance of electric shock in the presence of a partner than when alone.

It was hypothesized that the five-year-old children in the present study would be more willing to attempt the difficult when placed in pairs than when alone or before a large group (a class of 22 peers).

B. PROCEDURE

1. *Subjects*

The subjects were 44 five-year-olds enrolled in an experimental preprimary educational program following the Creative-Aesthetic approach (7, 8). The subjects were equally divided into two classes each taught by a master teacher assisted by a part-time teaching aide. The classes were integrated and roughly represented a stratified sample in a medium-sized Southern town. A deliberate attempt had been made to include some disadvantaged children in each class. While the majority of children in the two classes studied are white, almost all of the remaining children in the school are black.

2. *The Setting*

Essentially the same behavioral settings were maintained in both classrooms. These behavioral settings were deliberately designed to encourage experimentation, manipulation, sociodramatic play, construction, and creative expression. The experimental materials (bean bags and waste baskets) used in this study had been available in both classrooms from the beginning of the school year but had been used almost entirely during free activity periods. The author was accustomed to working with each class for about one hour each week. Early in the school year during a free activity period, the author had encouraged some of the children to try to throw the bean bags into the basket from different distances. No systematic work had been done either to teach the children to estimate their limits or to encourage them to try the task at distances difficult for them to succeed. Any such learning prior to the present study was incidental.

3. *Experimental Material*

The training and test task used in this study was adapted from Starkweather's (3) test of willingness to try difficult tasks. Starkweather's test apparatus consists of a target game in which a ball is rolled from varying distances to a target enclosed in a box. A marker consisting of a strip of cloth calibrated for distance is used in determining what is difficult for each child and in setting the easy and difficult task. Prior to the development of this instrument, Starkweather (3) had used a bean bag target game as a test of willingness to try difficult tasks but had abandoned it.

In the present study, the experimental equipment consisted of two sets of three 12-ounce bean bags each, a marker designating five levels of difficulty, and a waste basket. The equipment used in the study was quite similar to that which had been available in the classrooms from the beginning of the school year. However, no marker had been used prior to this time.

4. *Orientation to the Task*

The target game was introduced as a part of the author's usual work with the children and was preceded by the reading and discussion of Mary McBurney Green's *Is It Hard? Is It Easy?* (1). As the book was read, the author and the children talked about whether it was easy for each of them to skip, tie shoelaces, climb trees, catch and throw a ball, take a bath, hold tickly worms, whistle, somersault, and pound nails so they go in straight. They were told that they could make the bean bag target game easy or difficult by standing close or far away from the target. In all three conditions, children were given opportunities to practice the task to "warm up" and to find out what was hard and easy for them.

5. *Peer Conditions*

Under all three conditions, each subject was given a practice session, an explanation of the scoring system (5 points for hitting the target from E, 4 points for hitting it from D, 3 points for hitting it from C, etc.), and three trials to hit the target.

For Conditions A and B, children were called from the classroom in random pairs to play the target game in the wide hallway just outside the classroom. Two sets of targets and markers were set up in such a way that the players were back to back and could not see what the other was doing. Under Condition A, the children practiced and played the game in pairs; under Condition B, they practiced and played alone. Under Condition C, they performed before the entire class and each class competed with the other for the ownership of a large ball unlike any available to them. After each child had performed, he was allowed to add whatever points he scored to the class score being kept on the blackboard. (The two classes achieved identical scores.) This procedure netted 22 subjects each in Conditions A and B and 44 subjects in Condition C. The 44 children in Condition C were the same ones who were in Conditions A and B. The study was conducted after the children had been in the preprimary program for about five months.

C. Results

A rough test of the hypothesis already stated can be obtained by comparing the proportion of easiest tasks chosen under each condition. These percentages were as follows: Condition A (Pairs), 13.7 percent; Condition B (Alone), 36.5 percent; Condition C (Class), 86.5 percent. In all cases, the differences in proportions are statistically significant at the one percent level of confidence. When working in pairs, children least frequently tried the target from the easiest position. When performing before the class, they were most conservative and most frequently tried the target from the easiest position.

A somewhat more precise test can be made by obtaining a distance score for trial, adding the distance score for the three trials by each child, and comparing the differences in means for the three conditions. The means and standard deviations, respectively, for the three conditions are as follows: Condition A (Pairs) 12.1, 3.3; Condition B (Alone), 6.7, 3.9; Condition C (Class), 3.8, 1.6. An analysis of variance yielded an F ratio of 64.64, significant at better than the .01 level. A t ratio of 4.61 was obtained for the difference in means for Conditions A and B and one of 1.91 was obtained for the difference between Conditions B and C, the former being significant at better than the .01 level and the latter being significant at about the .05 level.

D. Discussion

These results suggest that young children tend to be most willing to try difficult tasks when working in pairs and least willing to do so when performing before the entire class. In individual situations, they seem to be freer to adjust the difficulty of the task to their skill level than under the "pair" and "class" conditions.

One might question the extent to which one can generalize from the limited number of subjects included in this study. One might also wonder if the Creative-Aesthetic orientation of the program might not have influenced the results. Under the Creative-Aesthetic approach, there is constant emphasis on attempting difficult tasks and much effort is expended in developing a psychologically safe environment. Thus, it would be expected that children under the Creative-Aesthetic approach would feel freer than under other conditions to make errors or fail in the presence of the group. In both classes, however, the children were extremely reluctant to try the difficult before the class. Confidence in the results is strengthened by the fact that they were quite consistent within treatments and the differences between treatments were quite strong.

From these results, it would be expected that having children work on school tasks in pairs would influence them to try difficult tasks, while requiring them to perform before the entire class would discourage them from doing so. In the latter condition, none of the children in either class tried to hit the target from the five-point line and only two of them attempted to hit it from the four- or three-point line. When working in pairs, only three of the 22 children tried to hit the target from the easiest position; most of them held doggedly to the four- and five-point lines.

E. SUMMARY

In order to learn and to grow, children have to be willing to attempt difficult tasks. To facilitate learning and growth, teachers should be able to create the social or peer conditions conducive to attempting tasks of appropriate difficulty.

It was hypothesized that five-year-old children in an experimental preprimary school program would be more willing to attempt difficult tasks when placed in pairs than when alone or before their entire class.

The task used in this study was a target game in which the children attempted to throw bean bags into a basket from distances that made the task either easy or difficult, depending upon the choice of the child. In Condition A, the children played the game in pairs; in Condition B, alone; and in Condition C, before the entire class. There were 22 children each in Conditions A and B and 44 in Condition C.

The results clearly supported the hypothesis and further suggested that children are least willing to attempt the difficult when performing before the entire class.

REFERENCES

1. GREEN, M. M. Is It Hard? Is It Easy? New York: Young Scott Books, 1960.
2. MILLER, I., BENSEN, S. B., SEIDMAN, D., & MEELAND, T. Influence of a partner on tolerance for a self-administered electric shock. Paper presented at annual meeting of Western Psychological Association, Berkeley, California, March 29, 1956.
3. STARKWEATHER, E. K. Preschool children's willingness to try difficult tasks. Final Report on Cooperative Research Project No. 5-0333, Oklahoma State University, Stillwater, Oklahoma, 1966.
4. TORRANCE, E. P. Different ways of learning for different kinds of children. In E. P. Torrance & R. D. Strom (Eds.), *Mental Health and Achievement.* New York: Wiley, 1965. Pp. 253-262.
5. ———. Does evaluative feedback facilitate creative thinking? Newsletter, Division 2 of the American Psychological Association, Spring-Summer, 1966.
6. ———. Comparative studies of stress-seeking in the imaginative stories of preadolescents in twelve different subcultures. In S. Z. Klausner (Ed.), *Why Man Takes Chances.* Garden City, N.Y.: Doubleday, 1968. Pp. 195-233.
7. TORRANCE, E. P., & FORTSON, L. R. Creativity among young children and the creative-aesthetic approach. *Educ.,* 1968, **89**, 27-30.
8. TORRANCE, E. P., FORTSON, L. R., & DIENER, C. Creative-aesthetic way of developing intellectual skills among five-year olds. *J. Res. & Devel. in Educ.,* 1968, **1**(3), 58-69.

The University of Georgia
College of Education
Athens, Georgia 30601

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 195-197.

THE USE OF MATRIX RATHER THAN SERIAL PRESENTATIONS OF NONSENSE SYLLABLES*

University of Windsor, Ontario

ALAN KRICHEV AND BRENTON HAZLEWOOD

A. INTRODUCTION

Most experiments requiring the rote memorization of nonsense syllables or similar material have presented such material in a sequential (serial) list form. This helps to organize the material, allows the formation of associations between items and is the method typically employed when one is faced with the task of memorizing written material.

However, for a given experiment such a presentation can have drawbacks. From a practical standpoint, it can be difficult to present a long list, especially by the use of a slide or overhead projector to groups of *S*s. In addition, there are two disadvantages which may influence the results: (*a*) when a considerable amount of low-meaningfulness material needs to be learned, the monotony of the task may influence the results, and (*b*) it is a task which is familiar to most *S*s and, therefore, may cause them to lose interest quickly.

Such experimenters as Anderson and Ross (1), Hebb (2), and Woodworth (3) have presented material to be memorized in a matrix form (e.g., five rows of five items each). This method has the advantage of saving space and of being different from the usual method of organizing material to be memorized. At the same time, it could be used to vary the task when considerable memorization is needed.

It was the purpose of the present paper to compare the matrix with the sequential list presentation methods to ascertain whether the methods could be used interchangeably.

B. METHOD

1. *Subjects*

The *S*s were 60 University of Windsor undergraduates enrolled in an introductory psychology course. There were 32 men and 28 women. Eight men and seven women were randomly assigned to each of four groups: serial list-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on April 23, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

position (SP), serial list-random (SR), matrix-position (MP), and matrix-random (MR).

2. *Procedure*

Testing was done on the *S*s in groups of five, all of whom were from the same experimental group. They were brought into a classroom which had a movie screen in the front and were seated 15 feet from the screen and facing it. Each group of five was shown nine CVC trigrams on the screen. Groups SP and SR were shown the trigrams arranged in a serial list in the standard fashion, although the list was unnumbered; Groups MP and MR were shown them arranged in a 3 × 3 matrix. Each *S* was given 15 prenumbered answer sheets with either nine lines arranged vertically or a blank matrix.

Groups SP and MP were told to write down each trigram they could remember in exactly the same position or square as it appeared on the screen. Groups SR and MR were told to write down each trigram they could remember in any position or square, regardless of where it appeared on the screen.

*S*s received a total of 15 trials each. Each trial consisted of a five-second presentation period and a 30-second recall period. There were five seconds between trials to allow the *S*s to turn over their answer sheet so they would have no knowledge of results readily available. *S*s repeated the procedure for all 15 trials, even if they recalled all nine trigrams correctly on any given trial.

C. RESULTS

Table 1 presents the percentage of correct responses for each group at the end of each block of three trials. Analysis of variance indicated that neither the type of group ($F = 2.24$; $df = 3$; $p > .05$) nor the group by trial interaction ($F = 1.07$; $df = 12$; $p > .05$) had a significant effect on the amount of material learned. The sex of the *S*s and whether they had to respond by exact position also had no effect.

TABLE 1
NUMBER OF CORRECT RESPONSES PER BLOCK OF THREE TRIALS

Group	1-3	4-6	Trials 7-9	10-12	13-15	Total correct	% correct
SP	123	251	323	370	391	1458	72
SR	120	211	264	315	322	1232	61
MP	130	279	343	376	386	1514	75
MR	134	256	318	350	367	1425	70

D. CONCLUSION

The present results indicate that there is no essential difference between the matrix and the sequential list methods of stimulus presentation. Hence, the choice of the matrix form is indicated when the technical issues of limited presentation space or task variability require a different or more compact form.

REFERENCES

1. ANDERSON, S. B., & ROSS, S. Memory for items in a matrix. *Amer. J. Psychol.,* 1955, **68**, 595-604.
2. HEBB, D. O. The Organization of Behavior. New York: Wiley, 1949.
3. WOODWORTH, R. S. Experimental Psychology. New York: Holt, 1938.

Department of Psychology
University of Windsor
Windsor, Ontario
Canada

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 199-209.

INTERSTRESSOR CONSISTENCY OF SUBJECT DIFFERENCES IN AUTONOMIC ACTIVITY*[1]

U.S. Army Research Institute of Environmental Medicine, Natick, Massachusetts

GARY W. R. PATTON

A. INTRODUCTION

The purpose of this paper was to determine whether there are consistent differences among *S*s in the character of their autonomic nervous system (ANS) response to varied thermal and psychological stresses. The prediction was made that, regardless of the exact ANS response pattern elicited by the stress (2), (*a*) *S*s would respond consistently; that is, an *S* would exhibit a characteristic amount of sympathetic nervous system (SNS) activity in a particular measure under a variety of stresses, and (*b*) individuals classed as sympathetic (12) under rest conditions would display relatively more SNS activity in all autonomic measures during each stress than *S*s characterized as parasympathetic. In terms of measures employed here, greater SNS activity was indicated by increasing pulse rates, systolic blood pressures, palmar skin conductances, and decreasing instep skin temperatures (5).

Lacey and Lacey (7) showed that consistent intersubject differences occur for any given measure; however, they did not find that one *S* consistently exceeded another in a number of simultaneously-recorded SNS measures under varying stress conditions. In contrast, Jones (3) and Jost, Ruilmann, Hill, and Gulo (4) demonstrated that certain patient groups, such as hypertensives and neurotics, consistently display more SNS activity than normal *S*s in a number of concurrently-recorded autonomic measures under a variety of nonthermal stresses. The present study extended and elaborated previous work in the following ways: (*a*) Since neurotic patients consistently exhibit more SNS activity in a number of measures than normal individuals, an attempt was made to determine whether normal individuals can likewise be differentiated in terms of their SNS responses to stress. (*b*) The causes of any significant differences in SNS activity between sympathetic and parasympathetic *S*s

---

* Received in the Editorial Office, Provincetown, Massachusetts, on April 23, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The author gratefully acknowledges the assistance of Mr. James Henderson, Miss Elaine Olshan, Mr. Robert Thompson, and Mr. Robert Wiesen in the collection, reduction, and analysis of the data.

during stress were investigated. Such differences may appear either because sympathetic *S*s characteristically display more SNS activity than parasympathetic *S*s under nonstress, as well as stress conditions (i.e., they are basically "biased" toward greater SNS activity under any circumstance), or because sympathetic *S*s show greater increases in SNS activity when confronted with a stress. Efforts were made to determine whether either or both of these factors were operative. (*c*) Cold and heat exposures were included among the four stress conditions to determine if *S* differences recorded during psychological stresses are maintained during thermal stresses. As noted by Ruch, Patton, Woodbury, and Towe (10), thermoregulatory activity, like response to psychological stress, is mediated by the ANS; therefore, if an *S* is hyperreactive to psychological stress, he may likewise be highly responsive to thermal stresses.

B. METHOD

1. *Subjects*

Twenty-four U.S. Army enlisted men between the ages of 19 and 26 served as *S*s in this experiment. Prior to the study, autonomic balance scores (12) were recorded for each *S*. Based on these results, *S*s were divided into two equal groups consisting of (*a*) the 12 *S*s with the lowest (most sympathetic) and (*b*) the 12 *S*s with the highest (most parasympathetic) scores. The mean and standard deviation for *S*s placed in the sympathetic category were 68.66 and 2.58, respectively; the corresponding values for *S*s assigned to the parasympathetic group were 78.34 and 5.39. After corrections for heterogeneity of variance were made (1), a t test indicated that these two groups differed significantly in autonomic balance ($t = 5.61$, $p < .001$).

2. *Experimental Conditions*

Each *S* was individually subjected to the following four stress conditions: (*a*) *S* was exposed to a temperature of 50±1°F. with 50 percent R.H. and a wind speed of less than two mph. for 60 minutes. (*b*) *S* was exposed to a temperature of 110±1°F. with 20 percent R.H. and a wind speed less than two mph. for 60 minutes. (*c*) *S* was shown two technicolor movies of surgery entitled "Debridement of Wounds of the Extremities" and "Management and Technic of Pulmonary Lobectomy." (*d*) *S* was asked to solve 35 five- and six-letter anagrams. The 80-second time limit for solution of each anagram was short enough that every *S* failed to solve at least one-half of the anagrams. A fifth session in which *S* watched a comedy movie provided nonstress comparison values for later statistical analyses.

3. *Experimental Design*

Each *S* underwent one experimental condition per day for five consecutive days, each session lasting 60 minutes. Conditions were administered on different days to prevent the artifactual appearance of autonomic response constancy based on close proximity of conditions in time, a problem noted by Wenger, Clemens, Coleman, Cullen, and Engel (13). Likewise, the testing time for each *S* was randomly varied from day to day to eliminate consistency in autonomic response based on similarities in diurnal cycle.

All five conditions could not be counterbalanced because of difficulty in scheduling climatic chambers for heat and cold exposures at appropriate times. However, a partially counterbalanced schedule was employed whereby all *S*s were given the heat exposure on the first day of the experiment, the cold exposure on the fifth or last day, and on the three intervening days, the three remaining conditions were counterbalanced among *S*s for order and sequence. On those days when the *S* was not heat- or cold-exposed, room temperature was maintained at 80±2°F. with 50 percent R.H. and less than two mph. wind speed.

4. *Procedure*

To eliminate post-prandial effects, *S*s were not allowed to eat in the two hours preceding testing. Each day, upon arriving at the laboratory, *S*, who was dressed only in shorts, was placed in a semireclining position in a chair and then outfitted with equipment to record pulse rate, systolic blood pressure, palmar skin conductance, and instep temperature. The first three measures were recorded on a Grass Model 7 polygraph, while the fourth was read by a Yellow Springs Model 47 tele-thermometer.

To record pulse rate, a Grass PTTI photoelectric pulse transducer was placed on the third finger of the *S*'s left hand. The transduced signal was then fed back to a Grass Model 7P4A tachograph preamplifier on the polygraph. A 30-second sample of each minute's pulse rate recording formed the basis for subsequent data reduction and analysis.

A combination of a Grass Model 7P8B sphygmomanometer and E & M automatic cuff pump was used to measure systolic blood pressure at two-minute intervals. The microphone from the sphygmomanometer preamplifier was placed over the brachial artery at the antecubital fossa of the *S*'s left arm to record the Korotkov sounds.

Palmar skin conductance was measured by passing a 50 μamp. current between two .80-inch diameter zinc chamber electrodes, which were filled with zinc sulfate electrode paste (9) and placed on the left palm and dorsal surface of the left wrist. This signal, originally recorded in ohms, was then fed to a Grass Model 7P1A low-level DC preamplifier on the polygraph. The final record was sampled every 60 seconds, and the recorded values were converted from ohms resistance to square root of the conductance (14) for subsequent data analyses.

Finally, instep temperature was monitored for 20 seconds every four minutes by a combination of a Yellow Springs Model 421 surface temperature thermistor, which was placed on the instep of the *S*'s right foot, and a Yellow Springs Model 47 tele-thermometer.

C. Results

1. *Individual Measures During the Nonstress Condition*

To determine if sympathetic *S*s displayed significantly more SNS activity than parasympathetic *S*s during nonstress conditions, the mean absolute values of pulse rate, systolic blood pressure, palmar conductance, and instep temperature recorded from these two groups during the comedy condition were contrasted by *t* tests. As shown in Table 1, sympathetic *S*s displayed more SNS activity in three of four measures; however, the differences between the groups were significant only for instep temperature.

2. *Composite Index of SNS Activity During the Nonstress Condition*

To gain a more general estimate of SNS activity during the nonstress condition, for each of the 24 *S*s, the mean absolute value for each measure was transformed into a T score to produce a distribution with a mean of 50, a standard deviation of 10, and a sample size of 24. The four T scores, each corresponding to one of the four autonomic measures taken from an *S* during one stress condition, were then added together. Finally, the mean composite score for the 12 sympathetic *S*s was compared with that for the 12 parasympathetic *S*s by means of a *t* test. As shown in Table 1, the composite index of SNS activity during the nonstress condition was significantly greater for sympathetic *S*s than for parasympathetic *S*s.

3. *Change in Individual Measures from Nonstress to Stress Conditions*

The next analysis was designed to determine (*a*) if *S*s display consistent increases in SNS activity with respect to their fellows under varying stress conditions and (*b*) if sympathetic *S*s display greater increases in SNS activity than parasympathetic *S*s. With the use of values recorded during the comedy condition as nonstress baseline figures, a measure of change, the au-

TABLE 1

MEAN AUTONOMIC RESPONSES OF SYMPATHETIC AND PARASYMPATHETIC *S*s DURING THE NONSTRESS CONDITION AND EACH OF THE FOUR STRESS CONDITIONS

Autonomic measures	*S* category	Nonstress condition		Stress conditions							
				Anagrams		Surgery		Cold		Heat	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pulse rate (beats/min)^{b}	Symp	67.75	7.06	69.92	4.37	69.06^{a}	6.43	70.68	8.98	82.19^{a}	6.03
	Para	63.92	6.76	68.00	7.35	63.73	6.19	67.48	7.80	74.24	10.11
Systolic blood pressure (mm Hg)	Symp	117.79	11.27	125.60	16.91	119.23	12.97	132.07^{d}	13.00	123.62	17.08
	Para	113.54	7.51	120.71	8.20	112.51	12.31	126.71	9.44	117.44	10.76
Square root palmar skin conductance (micromhos$^{1/2}$)^{b,c}	Symp	3.52	.64	4.06	1.26	4.67^{a}	.73	3.40	.91	6.29^{a}	2.04
	Para	3.54	.36	3.93	.78	3.64	.67	3.25	.36	4.51	1.23
Instep temperature (°F)	Symp	87.58^{a}	2.80	87.77	2.99	87.40	3.09	73.50	3.63	96.92	.82
	Para	89.38	1.29	87.99	2.03	89.00	2.98	74.44	1.78	96.86	.75
Composite response index (sum of T scores)^{b}	Symp	208.60^{a}	30.78	204.47	24.65	215.33^{a}	24.04	206.88	28.64	210.90^{a}	25.65
	Para	191.45	10.86	195.46	16.79	184.87	14.16	193.00	11.66	189.12	13.30

[a] p difference between *S* categories $< .05$ during this condition.
[b] p overall difference between *S* categories during stress conditions $< .05$ for this measure.
[c] p of *S* category × stress conditions interaction $< .05$ for this measure.
[d] $N = 11$.

tonomic lability score (6), was calculated for each of the S's four autonomic measures during each of the four stress conditions.

To determine if Ss display characteristic changes in a particular measure under varying stress conditions, the autonomic lability scores for the 24 Ss were ranked from 1 to 24 during each of the four stress conditions. The similarity of the rankings from one stress condition to the next was then tested by means of a Kendall coefficient of concordance, the preceding statistical analysis being completed once for each of the four autonomic measures. In all four cases, a significant tendency for interindividual changes in autonomic activity to remain stable from one stress to the next was noted ($W_{\text{systolic blood pressure}} = .54$, $p < .001$; $W_{\text{pulse rate}} = .48$, $p < .01$; $W_{\text{instep temperature}} = .47$, $p < .01$; $W_{\text{palmar conductance}} = .41$, $p < .05$).

The preceding analyses illustrated consistency of response change; however, they did not indicate whether sympathetic or parasympathetic Ss displayed greater increases in SNS activity. To answer the latter question, the autonomic lability scores for each of the four autonomic measures were subjected to analysis of variance (8). As shown in Table 2, with two exceptions, the sympathetic Ss were more reactive; however, only in the case of square root of palmar skin conductance was a significant difference, $F\ (1, 22) = 9.12$, $p < .01$, noted between groups. All four S category × condition interactions were nonsignificant, indicating that group differences were relatively consistent from one condition to the next.

4. *Composite Index of Change in SNS Activity from Nonstress to Stress Conditions*

In an analysis similar to one completed previously on the nonstress recordings, S's autonomic lability scores for all four physiological measures were added together to gain a more general index of change in autonomic activity during each stress condition. The summed scores attained by the 24 Ss were ranked from 1 to 24 during each stress condition, and the similarity of the rankings from one condition to the next was tested by the Kendall coefficient of concordance. The value thus produced was significant ($W = .55$, $p < .001$), indicating that S rankings in terms of overall autonomic change during stress remain fairly stable under varying stress conditions.

However, as shown in the bottom row of Table 2, differences between sympathetic and parasympathetic Ss in this measure did not reach significance when tested by analysis of variance. Again, the S category × condition interaction was nonsignificant.

TABLE 2
AUTONOMIC LABILITY SCORES OF SYMPATHETIC AND PARASYMPATHETIC *S*s DURING FOUR STRESS CONDITIONS

		Stress conditions							
		Anagrams		Surgery		Cold		Heat	
Autonomic measures	*S* category	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pulse rate	Symp	49.85	9.26	52.99	10.46	50.69	10.72	53.65	7.17
	Para	50.16	11.08	47.00	8.92	49.32	9.58	46.34	11.32
Systolic blood pressure	Symp	50.80	13.13	51.55	10.56	51.51[c]	10.47	51.62	12.25
	Para	49.20	5.91	48.46	9.61	48.50	9.78	48.40	8.16
Square root palmar skin conductance[b]	Symp	50.70	12.28	56.13[a]	8.51	51.57	11.72	54.89[a]	10.67
	Para	49.21	7.39	43.96	7.57	48.20	8.05	45.10	6.60
Instep temperature	Symp	48.23	11.39	50.53	11.04	50.44	13.22	50.08	10.51
	Para	51.78	8.51	49.49	9.30	49.57	5.79	49.95	9.92
Combined measure (sum of individual autonomic lability scores)	Symp	199.58	25.43	211.20	26.08	204.21	25.36	210.23	24.41
	Para	200.35	19.79	188.92	17.06	195.59	15.84	189.78	14.10

Note: All *S* category × stress condition interactions were nonsignificant.

[a] p difference between *S* categories $< .05$ during this condition.

[b] p overall difference between *S* categories during stress conditions $< .05$ for this measure.

[c] $N = 11$.

5. *Absolute Levels of Individual Measures During the Stress Conditions*

Stress effects can be gauged not only by noting the amount of change from nonstress to stress conditions, but also by measuring the magnitude of the resulting absolute levels of SNS activity. The following analyses were intended to indicate (*a*) if *S* rankings in terms of absolute autonomic activity remain stable under varying stress conditions, and (*b*) if sympathetic *S*s display consistently higher levels of SNS activity than parasympathetic *S*s.

As in preceding analyses, the Kendall coefficient of concordance was used to test the stability of *S* rankings of a particular measure under the four stress conditions. For all four measures, a significant tendency for interindividual differences in absolute autonomic activity to remain stable from one stress to the next was noted ($W_{\text{systolic blood pressure}} = .59$, $p < .001$; $W_{\text{pulse rate}} = .57$, $p < .001$; $W_{\text{palmar conductance}} = .47$, $p < .01$; $W_{\text{instep temperature}} = .46$, $p < .01$).

The consistency and extent of mean differences in the four autonomic measures between sympathetic and parasympathetic *S*s under the four stress conditions were tested by analysis of variance. The means in Table 1 show that, with the exception of instep temperature in the heat, sympathetic *S*s displayed more SNS activity than parasympathetic *S*s in every autonomic measure under every stress condition. The overall difference between the sympathetic and parasympathetic groups was significant for pulse rate, $F(1, 22) = 4.30$, $p < .05$), and square root of palmar skin conductance, $F(1, 22) = 7.83$, $p < .025$, but not for systolic blood pressure or instep temperature. In three of the four analyses, palmar skin conductance being the sole exception, nonsignificant *S* category × condition interactions were found indicating that, in these three cases, the magnitude of the differences between the sympathetic and parasympathetic groups remained relatively consistent from one stress to the next.

6. *Composite Index of Absolute Levels of SNS Activity During the Stress Conditions*

Composite indices of the absolute level of SNS activity displayed by each *S* during each of the four stress conditions were computed as previously described; that is, mean absolute values of each of the four autonomic measures recorded from an *S* during one session were converted to T scores, and the four T scores were added together. The *S*'s composite T scores were ranked during each condition, and, as in preceding sections, the similarity of the rankings from one stress condition to the next was tested by means of a Kendall coefficient of concordance. The resulting *W* was significant ($W = .62$,

$p < .001$), indicating that, relative to others, the total amount of SNS activity exhibited by an *S* tended to remain stable over stresses.

To determine if sympathetic *S*s consistently displayed more composite SNS activity than parasympathetic *S*s, the mean summated T scores for these two groups were compared over the four stress conditions by analysis of variance. As shown in the bottom row of Table 1, the sympathetic *S*s displayed more sympathetic activity, as evidenced by their significantly larger, $F\ (1, 22) = 7.52$, $p < .025$, overall scores. Again, in this analysis, the *S* category × condition interaction was nonsignificant, indicating that differences between the groups were fairly uniform from one stress condition to the next.

D. DISCUSSION

The preceding analyses indicate that, relative to other *S*s, *S*s respond consistently under varying stress conditions, both in terms of change in SNS activity from nonstress to stress conditions and in terms of absolute level of SNS activity displayed during stress. For example, an *S* who has both an above average increase in pulse and an above average pulse rate during one stress can be expected to behave similarly during other stress conditions. As noted by Sternbach (11), different stresses elicit different patterns of autonomic response. Thus, one stress may produce a higher mean pulse rate than another, and the hypothetical *S*'s pulse may rise and fall with the general mean. However, regardless of whether this *S*'s pulse rate is high or low in an absolute sense, during any given stress, it will always be above average in comparison with other *S*s' pulse rates.

Second, the results show that *S*s classed as sympathetic consistently exhibit more SNS activity during various stresses than do parasympathetic *S*s, both on an individual and a composite measure basis. Though the data are not conclusive, they tend to suggest that differences in SNS activity between sympathetic and parasympathetic *S*s during stress exist both because (*a*) sympathetic *S*s characteristically display higher levels of SNS activity during nonstress conditions (i.e., they have a head start when the stress is applied) and (*b*) sympathetic *S*s are more reactive to stresses.

Individual differences in learning (i.e., previous experiences with and attitude toward particular stresses) might logically be expected to modify constitutional differences in autonomic function among *S*s. For example, an *S* classed as sympathetic who has worked in a hospital might find movies of surgery less stressful and, as a result, display less sympathetic activity during the film than an *S* categorized as parasympathetic who has never seen an operation. Moreover, changes from "psychological" stress (the anagrams and films of surgery) to thermal stress (cold and heat exposures) might be expected to destroy the consistency of intersubject differences. Nevertheless, in this experiment, changes in the identity of the stress were not of sufficient importance to obliterate the basic differences in autonomic activity among *S*s, as witnessed by the consistent rankings of *S*s' autonomic measures from one stress condition to the next. Additional support for this statement can be found in Table 1 where four out of five relevant *S* category × condition interactions were found to be nonsignificant, indicating that, on the whole, response differences between sympathetic and parasympathetic *S*s remained fairly stable regardless of the condition being experienced.

E. Summary

The purpose of this study was to determine whether there are consistent differences among *S*s in the character of their autonomic nervous system (ANS) response to varied thermal and psychological stresses. The prediction was made that, regardless of the exact ANS response pattern elicited by the stress, (*a*) *S*s would respond consistently; that is, an *S* would exhibit a characteristic amount of sympathetic nervous system (SNS) activity in a particular measure under a variety of stressors and (*b*) individuals classed as sympathetic under rest conditions would display relatively more SNS activity during each stress than *S*s characterized as parasympathetic.

Twenty-four adult, male *S*s were classified as sympathetic or parasympathetic according to their autonomic balance scores and were then exposed to one nonstress (comedy movie) and four independent stress conditions (cold, heat, anagrams, and films of surgery) while recordings were taken of pulse rate, systolic blood pressure, palmar skin conductance, and instep temperature. Greater SNS activity was indicated by increasing pulse rates, systolic blood pressures, palmar skin conductances, and decreasing instep skin temperatures.

S rankings of (*a*) change in SNS activity from nonstress to stress conditions and (*b*) absolute level of SNS activity during stress were relatively consistent from one stress to the next regardless of the particular autonomic measure involved. Moreover, *S*s classed as sympathetic had consistently higher pulse rates and palmar skin conductances during all four stressors than those *S*s categorized as parasympathetic. Finally, when all four measures were combined into a more general index of SNS activity, the sympathetic *S*s had consistently higher combined scores. Differences in SNS activity between sympathetic and parasympathetic *S*s were tentatively attributed to (*a*) constitutional differences in autonomic activity and (*b*) the greater reactivity of sympathetic *S*s during stress.

REFERENCES

1. Cochran, W. G., & Cox, G. M. Experimental Designs. New York: Wiley, 1950.
2. Davis, R. C., Lundervold, A., & Miller, J. D. The pattern of somatic response during a repetitive motor task and its modification by visual stimuli. *J. Comp. Physiol. Psychol.*, 1957, **50**, 53-60.
3. Jones, M. Physiological response to stress in neurotic patients. *J. Ment. Sci.*, 1948, **94**, 392-427.
4. Jost, H., Ruilmann, C. J., Hill, T. S., & Gulo, M. J. Studies in hypertension: II. Central and autonomic nervous system reactions of hypertensive individuals to simple physical and psychologic stress situations. *J. Nerv. Ment. Dis.*, 1952, **115**, 152-162.
5. Kuntz, A. The Autonomic Nervous System. Philadelphia, Pa.: Lea & Febiger, 1953.
6. Lacey, J. I. The evaluation of autonomic responses: toward a general solution. *Ann. N.Y. Acad. Sci.*, 1956, **67**, 125-163.
7. Lacey, J. I., & Lacey, B. C. Verification and extension of the principle of autonomic response stereotypy. *Amer. J. Psychol.*, 1958, **71**, 50-73.
8. Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. Boston, Mass.: Houghton-Mifflin, 1956.
9. Lykken, D. T. Properties of electrodes used in electrodermal measurement. *J. Comp. Physiol. Psychol.*, 1959, **52**, 629-634.
10. Ruch, T. C., Patton, H. D., Woodbury, J. W., & Towe, A. L. Neurophysiology. Philadelphia, Pa.: Saunders, 1961.
11. Sternbach, R. A. Principles of Psychophysiology. New York: Academic, 1966.
12. Wenger, M. A. Studies of autonomic balance in army air force personnel. *Comp. Psychol. Monog.*, 1948, **19**, 1-111.
13. Wenger, M. A., Clemens, T. L., Coleman, D. R., Cullen, T. D., & Engel, B. T. Autonomic response specificity. *Psychosomat. Med.*, 1961, **23**, 185-193.
14. Woodworth, R. S., & Schlosberg, H. Experimental Psychology. New York: Holt, 1960.

United States Army Research Institute of
Environmental Medicine
United States Army Natick Laboratories
Natick, Massachusetts 01760

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 211-217.

MAGNITUDE SCALING OF FOODS*[1]

Department of Psychology, University of Rochester

JUDITH W. ONLEY, NINA SULLIVAN, AND SUSAN H. BORCYCK

What do we mean when we call something "rich" or "fattening"? The present study sought to evaluate the extent to which an individual can use the concepts "richness" and "fattening" in a quantitative manner. The problem was to gauge the extent to which various foods are perceived as "rich" or "fattening." The method of magnitude estimation, common to sensory psychophysics, though not widely applied to semantics, was used (6).

Questions of food preference and food scaling have concerned both food technologists and sensory psychologists in recent years. Harper gives references to the history of work on food preferences (3, 4); representative scaling studies include those by Bruvold and Gaffey (1) and Gregson and Russell (2). There has also been considerable practical interest in the derivation of adequate predictors of food consumption and preference [cf.: Pilgrim and Kamen (5)]. The present research has applicability in many of these areas, but was designed primarily to focus attention on the usefulness of a relatively new psychophysical methodology for a problem seemingly out of the realm of psychophysics—the evaluation of meaning, as applied to food evaluation.

Several group experiments were conducted; foods selected for investigation were selected from a variety of menus published in standard cookbooks. They are listed in Table 1, in one of the random orders in which they were presented. Subjects varied in age from 19 to 45, and included approximately equal sex distributions. Anyone who indicated prior knowledge of "Nutrition," or who had been in a special dietary regimen of any sort was excluded from participation. In each portion of the experiment, groups chosen at random within an Introductory Psychology class received a typewritten sheet (see appendix), on which food names were listed, with specific instructions concerning the judgment to be made, and the standard to be used. Instructions were similar for both "richness" and "fattening," and for the standards of "lamb chop," "ice cream," and "homogenized milk." Instructions followed the paradigm:

---

* Received in the Editorial Office, Provincetown, Massachusetts, on April 26, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This research was conducted with the assistance of a grant from the National Science Foundation for undergraduate research participation.

TABLE 1
PEARSON *r* CORRELATION COEFFICIENTS

Standards	Lamb chop	Ice cream	Homogenized milk
Richness *vs.* Calories	.52*	.52*	.72**
Fattening *vs.* Calories	.62**	.69**	.67**
Richness *vs.* Fattening	.60**	.86**	.92**

* Significant at the .05 level.
** Significant at the .01 level.

> Foods are often described as varying in their richness. Some foods are much richer than others, some not rich at all. Taking as a reference point, that we will call the richness of homogenized milk "100," indicate for each of the foods listed, how rich it is relative to this standard. If it seems half as rich, call it "50." If it seems three times as rich, call it 300 . . . and so forth. Assign each food on your list a number which indicates how rich you think it is, relative to the standard.

Whatever the standard, it was assigned the number "100." Separate groups were used in judgments of richness and of fattening, for each of three standards. Thus, the experiments involved 2 categories × 3 standards × $N = 19$, or a total of 114 subjects. A list of the specific foods judged is appended.

The basic data of these experiments are series of numerical estimates of the magnitude of "richness" and "fatteningness" of foods relative to the three standards. Initially, the data were analyzed in terms of major nutrient components (protein, carbohydrate, fat), as well as the obvious psychophysical correlate, caloric content. Nutrient classes provide less information than an-

APPENDIX

LIST OF ITEMS JUDGED FOR "RICHNESS" OR "FATTENING"

Plain pizza
Lobster in shell
Skim milk
Pecan pie
Blue cheese with crackers
Mashed potatoes
Dry toast
Milkshake
Mushrooms
Cottage cheese
Watermelon
Pork chop
Tea
Avocado pear
Tuna fish with noodles
Homogenized milk
Roast lamb
Cheeseburger
Eggplant
Pie à la mode
Baked ham
Chocolate eclair
Calves liver
Shrimp cocktail
Ice cream
Oysters on half shell
Chocolate layer cake
Pureed spinach
Soda crackers
Sauerkraut
Eggnog
Sweet roll
Prune whip
Plum pudding with sauce
Asparagus
Sponge cake
Lamb chop
Peanut butter sandwich
Cornish game hen
Crepes suzettes
Lemon sherbet
Broiled swordfish
Canned pears
Roast beef
Fresh strawberries
Hot chocolate
Pancakes with syrup

alysis based on caloric content, thence the latter analysis was pursued further. Table 1 summarizes the Pearson *r* correlation coefficients for this set of experiments; whatever the standard, we find reasonable and significant correlation between either "richness" or "fattening" and caloric content, adjudged

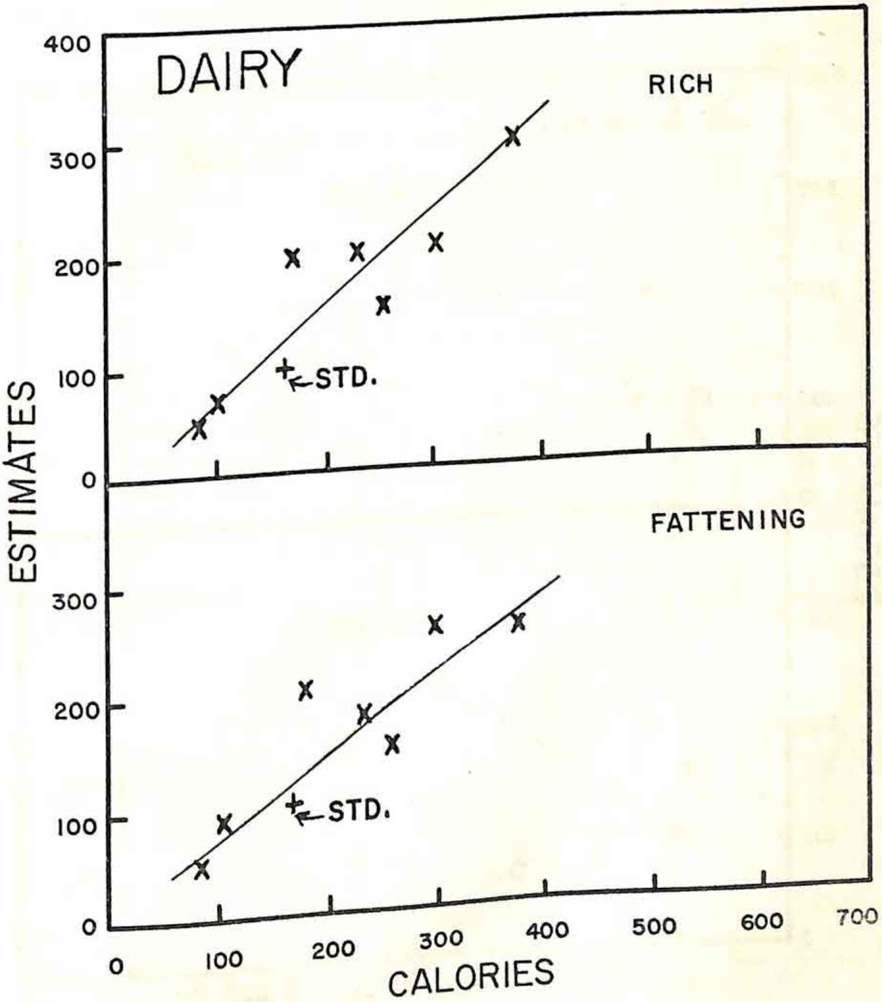


FIGURE 1
ESTIMATES OF RICH AND FATTENING FOR DAIRY PRODUCTS,
STANDARD IS HOMOGENIZED MILK CALLED "100"

for our stimuli via a standard cookbook (7). Richness and fattening judgments are highly correlated.

From the work of Pilgrim and Kamen (5), it is suggested that knowledge of food classes may provide predictive information with regard to food selection. Similarly, knowledge of food classes has been found to provide information which considerably improves the predictive value of magnitude functions

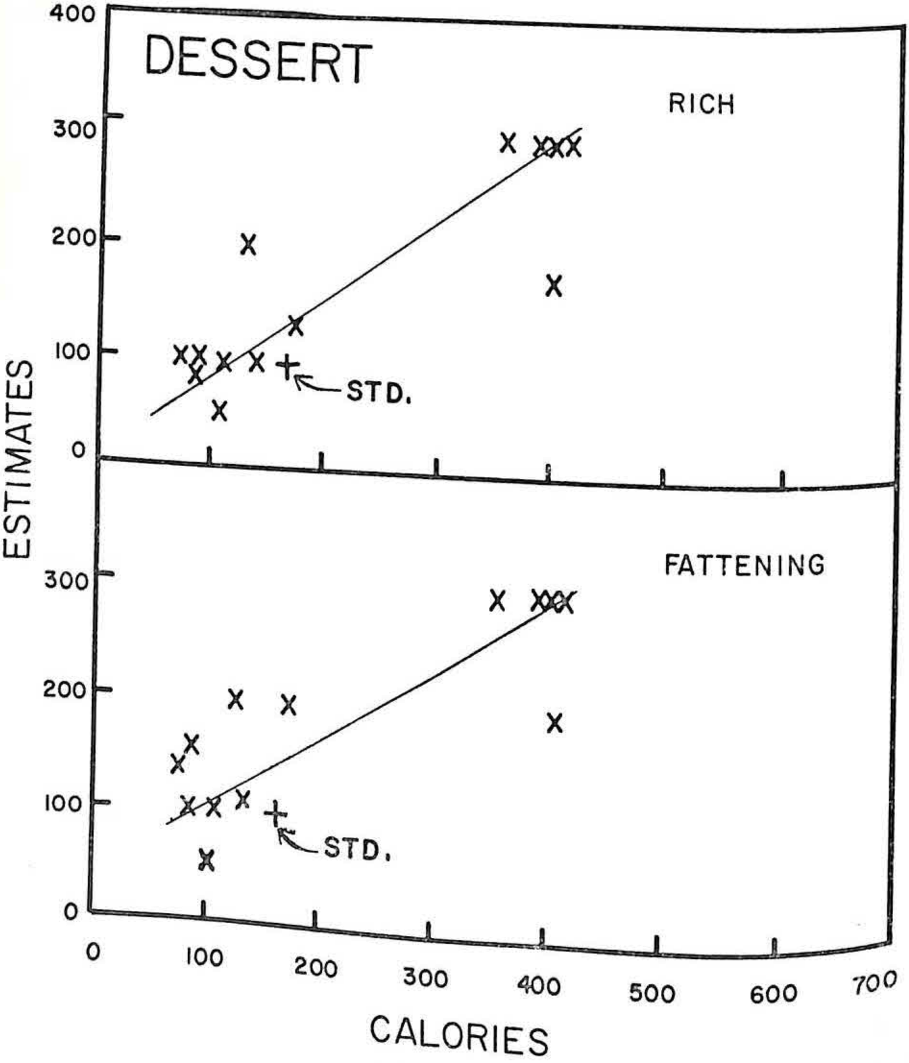


FIGURE 2
Similar Estimates for Desserts

relating either "richness" or "fattening" to caloric value. Dairy products, meat products, and desserts have been selected as useful categories for this particular sample of foods: in Figures 1, 2, and 3 magnitude estimates are shown for these three classes of foods, and the homogenized milk standard (which showed the highest general correlation with caloric content). Note that when stimuli are separated into food classes, high correlations exist for

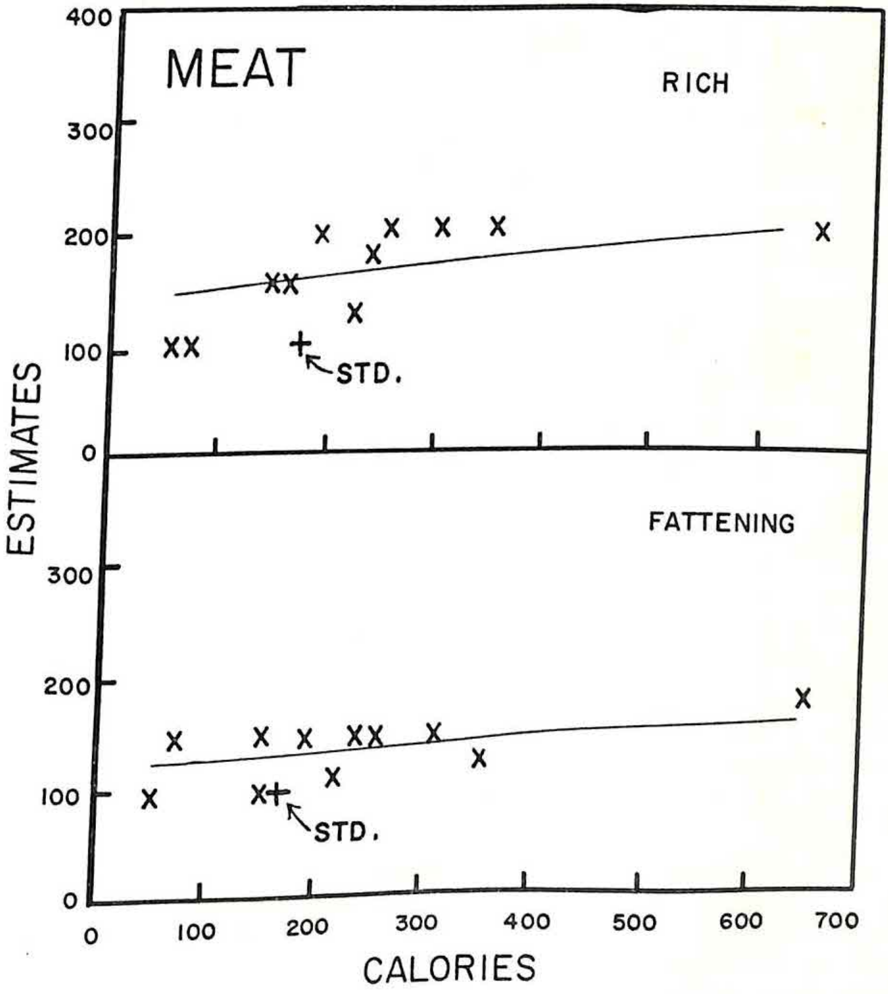


FIGURE 3
Similar Estimates for Meats

individual classes and that these, although all positive, differ significantly in magnitude for the three classes. The most obvious "maverick" is the class of meats; for these data, neither the term "rich" nor "fattening" is a good discriminator. Analogous results apply to judgments against the other two standards.

TABLE 2
"POWER LAW" EXPONENTS

Standards	Lamb chop	Ice cream	Homogenized milk
Food Types			
Milk Products			
Richness	1.54	1.59	1.15
Fattening	1.45	1.36	1.14
Meats			
Richness	.04	.15	.27
Fattening	.04	.20	.21
Desserts			
Richness	.96	.80	.92
Fattening	.89	.77	.95

Data of magnitude estimation studies in the realm of taste, as well as other sensory dimensions, are widely interpreted in terms of a power law form [note for example, 1, 2]. In the power law context, with the present data in log-log form, we find the exponents indicated in Table 2. Exponents for individual food types are highly similar for all three standards, but differ widely across food types. Again, there is evidence that caloric content is not a predictor of "rich" or "fattening," for the class labelled "meat products."

Is there any meaningful relation between what one calls "rich" and what is construed to be "fattening"? Present studies suggest an affirmative answer. Further, it appears that in at least one case, subjective estimates of "richness" bear little relation to caloric content. The term rich, as applied to foods, seems to lack "meaning," when applied to certain classes of foods; as applied to other classes, it has a definitive meaning.

While it is not clear that subjects base their judgment of the richness of foods or "fattening," on caloric values, the two types of judgment, for matched groups, are in themselves highly correlated. Foods that are judged to be rich, also tend to be judged to be fattening, and the reverse. The psychophysical method of magnitude estimation provides useful indices of meaning, with regard to the concepts of "richness" and "fattening," as applied to commonly experienced foods, evaluated according to caloric content. Food classes (in this case, meats, milk products, desserts) are useful predictors of the richness and fattening magnitudes; analysis of magnitude estimations for several standards yields food scaling functions which may be described by power law relations, but which are highly dependent upon food class.

REFERENCES

1. BRUVOLD, W. H., & GAFFEY, W. R. Subjective intensity of mineral taste in water. *J. Exper. Psychol.*, 1965, **69**, 369-374.
2. GREGSON, R. A., & RUSSELL, P. N. Psychophysical power law exponent value for sucrose intensity. *Percept. & Motor Skills*, 1965, **20**, 294.
3. HARPER, R. Food assessment and food acceptance as a psychological theme. *Occup. Psychol.*, 1960, **31**, 233-240.
4. ———. Psychologists' role in food-acceptance research. *Food Technol.*, 1962, **16**, 70-73.
5. PILGRIM, F. J., & KAMEN, J. M. Predictors of human food consumption. *Science*, 1963, **139**, 501-502.
6. STEVENS, S. S. On the operation known as judgment. *Amer. Scientist*, 1966, **54**, 385-400.
7. TRUAX, C., *Ed.* Ladies Home Journal Cookbook. Garden City, New York: Curtis Publishing Co. with Doubleday, 1960.

Department of Psychology
University of Rochester
Rochester, New York 14627

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 219-231.

THE COLOR-WORD INTERFERENCE TEST: THE EFFECTS OF VARIED COLOR-WORD COMBINATIONS UPON VERBAL RESPONSE LATENCY*[1]

Department of Psychology, Vassar College

JOYCE L. SICHEL AND KENNETH A. CHANDLER[2]

A. THE PROBLEM

In the study of individual differences, much interest has been shown recently in tasks requiring differential attention; that is, *S*s performing under conditions of distraction or when competing response tendencies are particularly strong. Studies in this general area have utilized the phenomenon of "color-word" interference for such assessment. When *S*s are asked to name the color in which a color-word is printed, and that color is different from the meaning of the color-word itself (e.g., the word RED printed in the color green), they experience considerable difficulty. Stroop used a serial format of such stimuli which has become known as the Stroop Test (24). Since then, many investigators have employed this test, and variations of it, to study individual differences in perception (11, 25), personality (14, 16), cognitive style (1, 2, 5, 9, 10, 20), development (6, 17, 18), and drugs and psychopathology (3, 19, 22, 23, 26). Jensen presents a comprehensive review of the findings of such investigations (13).

The measure of interference used by most investigators has been the difference between the time taken to name the color of a number of color blots and the time taken to name the color of a similar number of incongruently colored color-words. Jensen reports that this may well be the most sensitive measure of the "interference effect" (13). Several investigators have used variations of such a time difference to reflect the interference phenomenon,

---

* Received in the Editorial Office, Provincetown, Massachusetts, on April 26, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This study was implemented in part by resources from the National Science Foundation Instructional Improvement Grant GY-1841. Presented in abridged form to the Eastern Psychological Association, Washington, D.C., April, 1968 (21).

[2] The authors wish to express their appreciation to Mr. Cecil Hilles and Mr. David Brown of the Davenport Photographic Supply Co., New Haven, Connecticut for their assistance in the preparation of the color slides used in this study.

Communications should be directed to Dr. Kenneth A. Chandler at Vassar College, Poughkeepsie, N.Y. Copies of the modified Color Interference Test are available at a minimum cost to defray their expense.

but most of these have been no more than algebraic equivalents of this basic measure (12, 25). Almost all of this work, however, has been based on the implicit assumption that the time difference between naming a color blot and a color-word simply sums when numbers of such stimuli are employed.

Few investigators have reported data concerned directly with the nature of the test stimuli themselves. Klein varied the level of meaning of Stroop-type stimuli (15) and Jensen suggested that differences in the sequence of response (13) may contribute to total variance. Dalrymple and Budayr reported that certain serial lists of Stroop stimuli produced greater difficulty than others (7). Fraser, using the conventional Stroop format, did not find any differences in the interference effect when using selected color-word pair combinations (8). The present authors view such conflicting findings as reflecting the confounding effects inherent in the conventional Stroop design. Thus it would appear that insufficient attention has been paid to the fact that the stimuli words appear in a context of similar stimuli where they may be perceived in groups, though by instructions they are to be named singly. The opinion expressed here is that the effects of such word grouping on the color naming response may vary as a function of (*a*) the particular word combination, (*b*) the color and brightness of the stimuli, (*c*) the interaction of color and word meaning, and (*d*) variations in perceptual span.

In order to examine more closely changes in verbal responding that might occur as a function of such variables, verbal response latencies were obtained to singly presented color and color-word stimuli. It was felt that such a measure would be more sensitive and reliable than the usual total-time measures employed by previous investigators and have the added advantage of providing information concerning differences in color-naming difficulty for the discrete items employed. Single presentations of such stimuli provided the means of eliminating context effects present in the conventional Stroop format and of controlling perceptual span.

The usual Stroop color-word card allows triads and tetrads of similar stimuli, as well as many other recurring types of stimulus combinations. It was felt that this nonsystematic contextual bias confounds the actual interference effect. In view of this, an important aspect of the present study was the examination of the effects of different types of combinations of adjacent color-words. Dalrymple and Budayr (7) felt that differences in color-word pairs would produce differences in interference effects that would be maximal for that sequence of color-words where the suppression of the word reading response tendency to the first word would increase its unavailability for the color-naming response to the second word: for example, the word GREEN$_b$

(printed in the color blue) followed by the word RED_g (printed in the color green). However, their method did not take into account the various factors suggested in the preceding paragraphs, and perhaps this accounts for some ambiguity in their findings. The present authors felt that a condition meeting their specifications but having the added requirement of suppressing the color-naming response one has just made to the first word, RED_g (printed in the color green) paired with GREEN_r (printed in the color red) for example, would produce even greater interference effects. Several differing types of word-pair combinations were evaluated in this study; the particular word-pair combinations used here were selected on the basis of what was felt to be differing strengths of word-meaning and color-naming competition.

Thus the present study utilizes verbal response latencies to singly presented (*a*) color blots, (*b*) congruently and incongruently colored color-words, (*c*) paired combinations of color blots, and (*d*) paired combinations of congruently and incongruently colored color-words. Examination of such factors was directed towards determining the optimal conditions for constructing a more sensitive and reliable color-word interference test.

B. Apparatus and Procedure

*S*s were twenty-four female undergraduates at Vassar College. All were enrolled in the introductory course and served as *S*s to satisfy a course requirement. They were screened to rule out visual deficiencies, language fluency differences as a function of marked cultural variance, and medical conditions felt to be related to the task demand.

*S*s were seated in a moderately lighted room two feet from a Sawyer back-projection screen. Thirty-five mm color slides were presented automatically on this screen (Sawyer Projector Model 600); their appearance activated, by means of photo-cell and switching devices, a Standard Electric timer accurate to .01 second. *E* controlled the rate of presentation of the slides. *S*s were fitted with a throat microphone in the area of the superior carotid triangle which, through a Grason-Stadler Voice Operated relay (model E7300-A), provided the means by which *S*s' verbal response would stop the timer. The time between stimulus onset and *S*s' response was taken as the measure of verbal response latency. Several pre-experimental trials were employed to familiarize *S* with the task and to adjust throat microphone and electronic equipment for maximum sensitivity.

Stimuli used in this experiment consisted of 48 35-mm color slides of the following (Cf. Figure 1): (*a*) Colored rectangles: XXX_r; XXX_b; XXX_g. Subscripts indicate color in which rectangles were printed, 3 slides. (*b*) Congruently colored color-words: RED$_{r}$; BLUE$_{b}$; and GREEN$_{g}$. Subscripts indicate color in which word is printed, 3 slides. (*c*) Incongruently colored color-words: RED$_{b,g}$; GREEN$_{r,b}$; and BLUE$_{r,g}$. Subscripts indicate color in which word is printed, 6 slides. (*d*) Pairs of colored rectangles: e.g. XXX$_{r}$-XXX$_{g}$; XXX$_{b}$-XXX$_{r}$; etc. Subscripts indicate color in which rectangles are printed, 6 slides. (*e*) Pairs of congruently and incongruently colored color-words: 30 slides for the various combinations: e.g. Congruently colored pairs: RED$_{r}$-BLUE$_{b}$; GREEN$_{g}$-RED$_{r}$; etc. and Incongruently colored pairs: RED$_{g}$-GREEN$_{r}$; BLUE$_{r}$-GREEN$_{b}$; etc. Subscripts indicate color in which word is printed. Examples of test stimuli are presented in Figure 1. Figure 2 presents the particular color-word pair combinations selected on the basis of expected differences in competing response tendency strengths.

	COLORED RECTANGLES	CONGRUENTLY COLORED COLOR-WORDS	INCONGRUENTLY COLORED COLOR-WORDS
SINGLE STIMULI	XXX$_{r}$	RED$_{r}$	RED$_{b}$
	XXX$_{g}$	BLUE$_{b}$	BLUE$_{g}$
	XXX$_{b}$	GREEN$_{g}$	GREEN$_{r}$
PAIRED STIMULI	XXX$_{r}$ XXX$_{g}$	RED$_{r}$ GREEN$_{g}$	GREEN$_{r}$ RED$_{g}$
	XXX$_{b}$ XXX$_{r}$	BLUE$_{b}$ RED$_{r}$	RED$_{g}$ BLUE$_{g}$
	XXX$_{g}$ XXX$_{b}$	GREEN$_{g}$ BLUE$_{b}$	BLUE$_{r}$ RED$_{g}$

*Subscripts indicate color in which stimulus is printed.

FIGURE 1
EXAMPLES OF TEST STIMULI USED IN COLOR NAMING EXPERIMENTS

*S*s' task was to name the color of the stimulus presented. Instructions to *S* were for single stimuli: You are to NAME THE COLOR of what you see on each slide as rapidly as you can. Some slides will have words on them, some just colored rectangles. IN ALL CASES you are to respond with the NAME OF THE COLOR; it will be red, blue or green. For paired stimuli: On some slides there will be more than one item.

In these cases you are to name the color of the item on the left first, then the one on the right.

For the single stimuli experiment a Latin square design was employed so that the effects of sequence and order of presentation could be evaluated; analysis of variance did not reveal significant order or sequence effects. Thus a 3 × 3 factorial analysis of variance design was used for the single stimuli studies and a 5 × 6 factorial design for the paired stimuli studies. Each *S* was tested under all conditions, four measures for each of the forty-eight test stimuli.

LEVELS OF INTERFERENCE

Minimal	Intermediate	Maximal
CONGRUENTLY COLORED COLOR WORDS	INCONGRUENTLY COLORED COLOR WORDS I	INCONGRUENTLY COLORED COLOR WORDS II
RED$_r$ BLUE$_b$	GREEN$_r$ GREEN$_b$	BLUE$_r$ GREEN$_b$
RED$_r$ GREEN$_g$	BLUE$_r$ RED$_g$	GREEN$_r$ RED$_g$
BLUE$_b$ RED$_r$	GREEN$_b$ GREEN$_r$	RED$_b$ GREEN$_r$
GREEN$_g$ BLUE$_b$	RED$_b$ BLUE$_g$	BLUE$_g$ GREEN$_b$

*Subscripts indicate color in which color-word is printed

FIGURE 2
EXAMPLES OF COLOR-WORD PAIRS VARYING IN EXPECTED LEVELS OF INTERFERENCE

C. RESULTS

The expectation that verbal response latencies would provide a highly reliable measure for color naming tasks is supported by the results revealed in this study. The odd-even reliabilities for the single stimuli was found to be +.66 and for the paired stimuli +.69; both significant beyond the .01 level of confidence. Reliabilities between the various test conditions ranged from +.88 to +.96 revealing highly stable individual performance. Retest of *S*s did not produce any significant differences in verbal response latencies to the color naming tasks. It was hypothesized that verbal response latencies would differ as a function of the variables presented below.

1. *The Effects of Differently Colored Color-Blots and Different Color-Words Congruently Colored*

That verbal response latencies do differ as a function of color for both color blots and different congruently colored color-words is seen in Table 1.

It can be seen that verbal response latencies to RED, BLUE, and GREEN rectangles are significantly different as are those for the congruently colored color-words, RED_r, BLUE_b, and GREEN_g. The hypothesis that such differences would occur is thus supported. It is interesting to note that in this comparison the red colored rectangle and the color-word RED printed in red both have the shortest response latencies whereas green in both instances has the longest latency. This is consistent with the findings of Chandler and Palubinskas (4) who report that across three differing levels of judged brightness red colored rectangles were responded to more quickly than green colored rectangles; those colored blue appear to be intermediate between red and green.

TABLE 1
VERBAL RESPONSE LATENCIES TO COLORED RECTANGLES AND CONGRUENTLY COLORED COLOR-WORDS

Variable[a]		Mean[b] (milliseconds)	SE_M	N
Colored rectangles	XXX_r	508	16	24
	XXX_b	535	19	24
	XXX_g	579	19	24
Congruently colored color-words	RED_r	594	16	24
	BLUE_b	643	18	24
	GREEN_g	698	20	24

[a] Subscripts indicate color in which stimulus is printed.
[b] Differences between means within rectangles and means within color-words are significant at or beyond the .01 level of confidence.

2. *Congruently Colored vs. Incongruently Colored Color-Words*

It is to be expected that incongruently colored color-words would reflect the interference effect through increased verbal latencies when compared to congruently colored color-words. That this was the case is shown in Table 2 and presented graphically in Figure 3. Here we see that in every instance the incongruent coloring of a color-word resulted in an increased verbal latency. The shortest latencies presented here occurred with those color-words colored red which again reflected the saliency of this particular hue.

3. *Singly Presented Color-Words vs. Color-Word Pairs*

It was anticipated that congruently colored and incongruently colored color-words would differ in terms of verbal latencies as a function of being presented as pairs (two color words) compared to single presentation (one color word). The expectation was that the presence of the second color-word in the color-word pair, regardless of its color or meaning, would increase competing response tendencies with a resultant increase in verbal response latency. The latency for singly presented color-words, whether colored congruently or not, was found to be 733 milliseconds, while that for the color-word pairs was 941 milliseconds. The hypothesis that competing response tendencies would be enhanced by the presence of a second color-word and this reflected in an increased verbal response latency is thus supported. The correlation between single color-word stimuli verbal response latencies and those obtained with paired combinations is +.98 reflecting the high degree of consistency of individual performance.

TABLE 2
VERBAL RESPONSE LATENCIES FOR CONGRUENTLY AND INCONGRUENTLY COLORED COLOR-WORDS[a]

Color-Word[c]	Mean (ms)[b]	SE_M
Congruently colored		
RED_r	594	16
BLUE_b	643	17
GREEN_g	698	20
Incongruently colored		
RED_b	810	33
RED_g	799	16
BLUE_r	711	23
BLUE_g	813	22
GREEN_r	723	26
GREEN_b	810	25

[a] Increase in verbal latency of incongruently colored color-words over congruently colored color-words was 34 percent for RED, 18 percent for BLUE, and 20 percent for GREEN.

[b] A difference of 44 ms is significant at or beyond the .05 level of confidence.

[c] Subscripts indicate the color in which the word is printed.

4. *Different Types of Incongruently Colored Color-Word Pairs*

The effects of differing types of color-word pair combinations used in this study (see Figure 2) were reflected in differing response latencies. These results are presented in Figure 4. The type of color-word pair combination expected to pose the most difficulty for *S*s, and hence produce the longest verbal response latency, did in fact do so. This was the condition, indicated in this figure by Incongruently Colored II, of color-word pairs where the reading response tendency to the first word was the same as the color-naming response to the second word and where the color-naming response to the first word had to be suppressed because it was the word reading response to the second word.

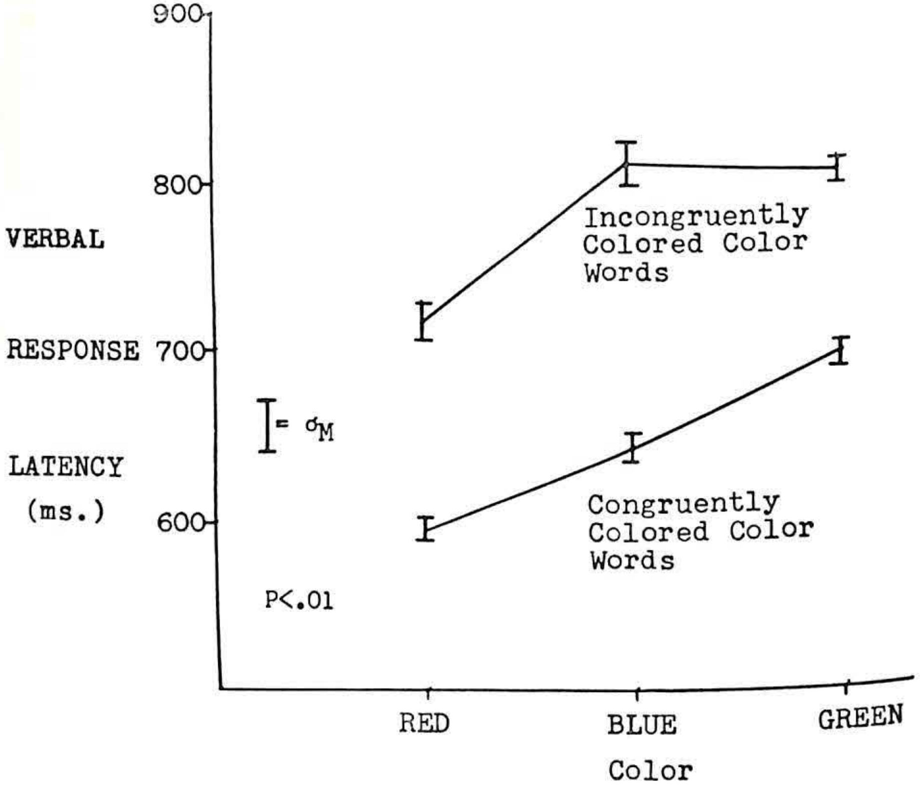


FIGURE 3
RESPONSE LATENCIES FOR INCONGRUENTLY AND CONGRUENTLY COLORED COLOR-WORDS

D. DISCUSSION

It is clear, first of all, that verbal response latencies differed as a function of differences in color and brightness of the stimuli. The findings reported here for these variables are also consistent with those of Chandler and Palubinskas (4). Both colored rectangles and color-words printed in colors congruent with their meanings showed shorter verbal response latencies when printed in red than for blue and green.

When color-words were presented in colors incongruent with their meanings, verbal response latencies increased over those for congruently colored color-words, thus demonstrating the basic color-word interference effect. The magnitude of this effect, however, was itself modified by the factors of hue and brightness and it is clear that this interference effect was not the same for all word by color combinations. A significant color by word interaction does occur.

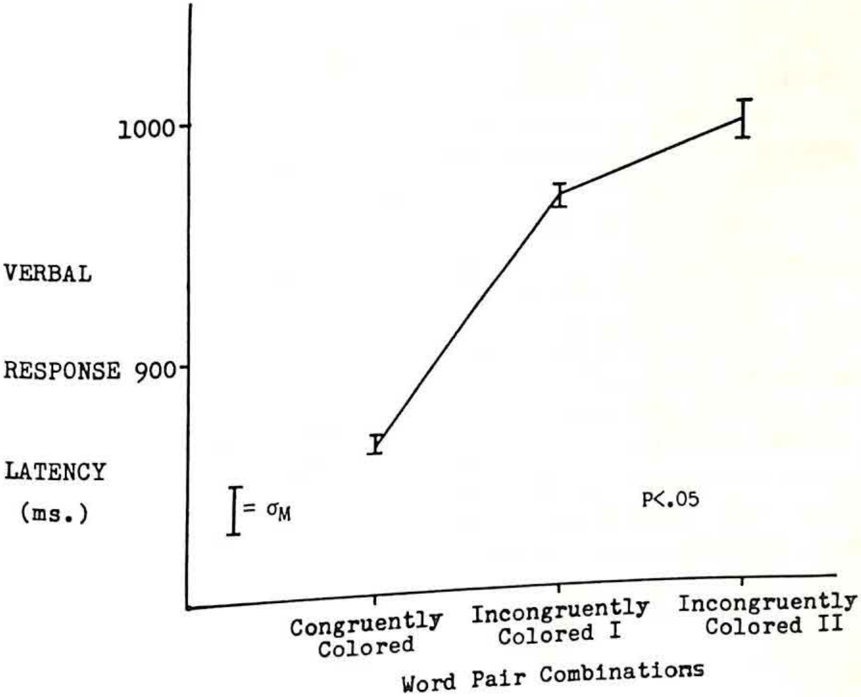


FIGURE 4
Verbal Response Latencies for Differently Colored Color-Word Pair Combinations

In addition to the factors of hue and brightness, it was also found that the extent and type of word context significantly affected the interference phenomenon. When word context was introduced, here by pairing color-words, an increased response latency was shown over that obtained for single color-words. It is apparent from these findings that the random selection of a specific color-word or pair of words could inadvertently operate either to enhance or to disrupt the strength and consistency of the interference effect. The conventional Stroop Test does in fact utilize randomly selected incongruently colored color-words massed together in as many varied combinations as might be expected from the randomization procedure (24). This confounds the specific relationships between color and word meaning with the effects of particular word-pair combinations. When the type of word-pair combinations is brought under more adequate control, the level of interference is enhanced and its consistency increased.

In order to examine such varying levels of interference, certain types of color-word combinations were selected. The work of Dalrymple and Budayr (7), as well as that of Fraser (8), suggested to the present authors that when suppression of a response tendency is greatest, subsequent interference should also be correspondingly maximal. Dalrymple and Budayr felt that the crucial suppression was that of the word reading response tendency to the first word which would result in an increase in its unavailability as the color-naming response to the second word: e.g., GREEN_b-RED_g; on the other hand, Fraser was of the opinion that the most crucial suppression was that of the color-naming response already uttered to the first word when it represented the incorrect word reading response to the second word: e.g., RED_g-GREEN_b. While their theoretical positions seem sound, their results were not altogether consistent, no doubt because they had little control of contextual confounding as expressed here. The present authors felt that a combination of both specifications should yield a condition of maximum difficulty in the color-naming task and therefore produce the greatest interference. If this were the case, then one could prepare a set of color-word stimuli in which maximum interference would occur and its effects *not* be confounded by other combinations that would tend to weaken it. Examples of such stimuli are presented in Figure 2 under the heading Incongruently Colored Color-Words II; their greater effect upon verbal latency is seen in Figure 4. This type of word-pair combination includes both mentioned above: those word-pairs requiring both sorts of suppression referred to earlier and found by the present authors to produce the highest degree of difficulty and also the stimuli requiring the suppression mentioned by Dalrymple and Budayr. It was found that the word-pair combinations suggested by Dalrymple and Budayr did produce the second greatest color-naming difficulty.

In consideration of the results reported in this study, particularly that verbal response latency differs as a function of type of color-word combination, it became clear that a modified color-word interference test was needed. On the basis of empirical findings, such a test was prepared and is presented in Figure 5. The particular color-word pairs were chosen in view of the fact that they did produce maximal interference effects and are also consistent with the theoretical framework discussed above. Each word-pair is to be read as a unit by itself with sufficient spacing between it and the next pair as the eye travels vertically down the stimulus plate. In this way confounding context effects are controlled, perceptual span held constant, and only those word-pair combinations are used that yield maximal interference. Five plates, each of 20 color-word pairs, provide for brevity of form, while also offering the means for evaluating intrasubject reliability.

PLATE I	PLATE II	PLATE III	PLATE IV	PLATE V
BLUE$_r$-GREEN$_b$	GREEN$_r$-BLUE$_g$	BLUE$_g$-RED$_b$	BLUE$_g$-GREEN$_b$	BLUE$_g$-RED$_b$
GREEN$_b$-RED$_g$	BLUE$_r$-RED$_b$	BLUE$_g$-GREEN$_b$	GREEN$_b$-RED$_g$	BLUE$_r$-GREEN$_b$
GREEN$_r$-RED$_g$	GREEN$_r$-BLUE$_g$	BLUE$_r$-RED$_b$	RED$_b$-BLUE$_r$	GREEN$_r$-BLUE$_g$
RED$_b$-GREEN$_r$	BLUE$_g$-RED$_b$	GREEN$_r$-BLUE$_g$	GREEN$_b$-BLUE$_g$	BLUE$_g$-GREEN$_b$
BLUE$_r$-RED$_b$	BLUE$_g$-RED$_b$	GREEN$_r$-RED$_g$	RED$_b$-GREEN$_r$	GREEN$_b$-RED$_g$
GREEN$_b$-RED$_g$	GREEN$_b$-BLUE$_g$	BLUE$_r$-GREEN$_b$	RED$_b$-BLUE$_r$	GREEN$_b$-RED$_g$
BLUE$_r$-RED$_b$	BLUE$_r$-GREEN$_b$	BLUE$_g$-RED$_b$	BLUE$_g$-RED$_b$	BLUE$_g$-RED$_b$
GREEN$_b$-BLUE$_g$	BLUE$_r$-GREEN$_b$	RED$_b$-GREEN$_r$	GREEN$_r$-BLUE$_g$	BLUE$_r$-GREEN$_b$
GREEN$_r$-BLUE$_g$	GREEN$_b$-RED$_g$	GREEN$_r$-RED$_g$	BLUE$_r$-RED$_b$	GREEN$_b$-BLUE$_g$
RED$_b$-BLUE$_r$	BLUE$_g$-GREEN$_b$	GREEN$_b$-RED$_g$	GREEN$_b$-BLUE$_g$	GREEN$_r$-RED$_g$
RED$_b$-GREEN$_r$	GREEN$_b$-RED$_g$	BLUE$_r$-RED$_b$	BLUE$_r$-GREEN$_b$	GREEN$_r$-BLUE$_g$
GREEN$_b$-BLUE$_g$	BLUE$_r$RED$_b$	RED$_b$-BLUE$_r$	RED$_b$-GREEN$_r$	BLUE$_g$-GREEN$_b$
BLUE$_g$-GREEN$_b$	RED$_b$-BLUE$_r$	RED$_b$-GREEN$_r$	GREEN$_b$-RED$_g$	RED$_b$-GREEN$_r$
BLUE$_g$-RED$_b$	RED$_b$-BLUE$_r$	BLUE$_g$-GREEN$_b$	BLUE$_g$-GREEN$_b$	GREEN$_b$-BLUE$_g$
RED$_b$-BLUE$_r$	BLUE$_g$-GREEN$_b$	GREEN$_r$-BLUE$_g$	BLUE$_r$-RED$_b$	RED$_b$-BLUE$_r$
BLUE$_r$-GREEN$_b$	RED$_b$-GREEN$_r$	GREEN$_b$-BLUE$_g$	GREEN$_r$-RED$_g$	BLUE$_r$-RED$_b$
BLUE$_g$-GREEN$_b$	GREEN$_r$-RED$_g$	GREEN$_b$-RED$_g$	GREEN$_r$-BLUE$_g$	GREEN$_r$-RED$_g$
BLUE$_g$-RED$_b$	GREEN$_r$-RED$_g$	RED$_b$-BLUE$_r$	GREEN$_r$-RED$_g$	RED$_b$-GREEN$_r$
GREEN$_r$-RED$_g$	GREEN$_b$-BLUE$_g$	GREEN$_b$-BLUE$_g$	BLUE$_g$-RED$_b$	BLUE$_r$-RED$_b$
GREEN$_r$-BLUE$_g$	RED$_b$-GREEN$_r$	BLUE$_r$-GREEN$_b$	BLUE$_r$-GREEN$_b$	RED$_b$-BLUE$_r$

*Subscripts indicate color in which color-word is printed.

FIGURE 5

INCONGRUENTLY COLORED COLOR-WORD PAIR COMBINATIONS WHICH PRODUCE MAXIMUM INTERFERENCE EFFECTS

In its present use, total time for reading each plate is combined to yield a total interference card time score which may be compared to a similar total time score for an identically arranged set of plates of colored rectangles. In this way, the present authors have found that attention to the major variables has been effected, resulting in a more sensitive and stable measure of color-word interference.

E. Summary and Conclusions

Single and paired Stroop-type color-word stimuli were presented in succession to *S*s by means of color slides. Verbal reaction times for color naming of these stimuli were measured. The factors of hue, brightness, extent of word context, and type of word-pair combination were studied with respect to their contribution to the basic color-word interference effect. It was found that selected types of color-word pairs increased the magnitude and stability of the interference effect. A more refined and internally consistent color-word naming test utilizing such pairs in a printed format is presented.

References

1. Broverman, D. Cognitive style and intra-individual variation in abilities. *J. Personal.*, 1960, **28**, 240-256.
2. Broverman, D., & Lazarus, R. Individual differences in task performance under conditions of cognitive interference. *J. Personal.*, 1958, **26**, 94-105.
3. Calloway, E. The influence of amobarbital (amylobarbitone) and methamphetamine on the focus of attention. *J. Ment. Sci.*, 1959, **105**, 382-392.
4. Chandler, K. A., & Palubinskas, A. The effects of hue and brightness on verbal response latency in a color naming task. Paper read before Mount Holyoke Intercollegiate Conference in Psychology, Mount Holyoke College, South Hadley, Massachusetts, May, 1968.
5. Chandler, K. A., & Sichel, J. L. Body movement: a study of imbalanced support in groups differing in equilibrating tendencies. Paper read before Eastern Psychological Association, Boston, Massachusetts, April, 1967.
6. Comalli, P., Jr., Wapner, S., & Werner, H. Interference effects of Stroop color-word test in childhood, adulthood, and aging. *J. Genet. Psychol.*, 1962, **100**, 47-53.
7. Dalrymple, A., & Budayr, B. Examination of some aspects of the Stroop color-word test. *Percept. & Motor Skills*, 1966, **23**, 1211-1214.
8. Fraser, J. Verbal satiation and antagonistic response tendencies. *McGill Undergraduate Proj. in Psychol.*, 1963, **12**, 32-38.
9. Gardner, R., Holsman, P., Klein, G., Linton, H., & Spence, D. Cognitive control: a study of individual consistencies in cognitive behavior. *Psychol. Issues*, 1959, **1**(4), 1-147.
10. Holt, R. Cognitive controls and primary processes. *J. Psychol. Res.*, 1960, **4**, 1-8.
11. Houston, B. K., & Jones, T. M. Distraction and Stroop color-word performance. *J. Exper. Psychol.*, 1967, **74**, 54-56.
12. Jensen, A. Scoring the Stroop test. *Acta Psychologica*, 1965, **24**, 398-408.
13. Jensen, A., & Rohwer, W., Jr. The Stroop color-word test: A review. *Acta Psychologica*, 1966, **25**(1), 36-93.
14. Klein, G. Need regulation. In M. Jones (Ed.), *Nebraska Symposium on Motivation*. Lincoln, Neb.: Univ. Nebraska Press, 1954. Pp. 224-274.
15. ———. Semantic power measured through the interference of words with color naming. *Amer. J. Psychol.*, 1964, **77**, 576-588.

16. LAZARUS, R., BAKER, R., BROVERMAN, D., & MAYER, J. Personality and psychological stress. *J. Personal.,* 1957, **25**, 559-577.

17. RAND, G., WAPNER, S., WERNER, H., & MCFARLAND, J. Age differences in performance on the Stroop color-word test. *J. Personal.,* 1963, **31**, 534-558.

18. SCHILLER, P. Developmental study of color-word interference. *J. Exper. Psychol.,* 1966, **72**(1), 105-108.

19. SCHWARTZ, M., & SHAGASS, C. Responses to colored and conflict inducing stimuli in a psychiatric population. *Percept. & Motor Skills,* 1965, **21**(2), 618.

20. SICHEL, J. L. An experimental study of interference phenomena in cognitive sets. Unpublished Master's thesis, Vassar College, Poughkeepsie, New York, 1967.

21. SICHEL, J. L., & CHANDLER, K. A. Response latency in color-word interference tasks. Paper read before Eastern Psychological Association, Washington, D.C., April, 1968.

22. SILVERSTEIN, A., & FRANKEN, R. Performance of the mentally retarded on the Stroop color-word test. *Percept. & Motor Skills,* 1960, **11**, 245-252.

23. SMITH, G., & NYMAN, G. Psychopathologic behavior in a serial experiment. *Lunds. U. Arrskr., Avd.* 2, 1959, **56**(5), 3-73.

24. STROOP, J. Studies of interference in serial verbal reactions. *J. Exper. Psychol.,* 1935, **18**, 643-661.

25. THURSTONE, L. A factorial study of perception. *Psychometric Monograph* (No. 4). Chicago, Ill.: University of Chicago Press, 1944.

26. WAPNER, S., & KRUS, D. Effects of lysergic acid diethylamide and differences between normals and schizophrenics on the Stroop color-word test. *J. Neuropsychiat.,* 1960, **2**, 76-81.

Department of Psychology
Vassar College
Poughkeepsie, New York 12601

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 233-242.

THE SELF-CONCEPT OF CREATIVE ADOLESCENTS*[1]

Department of Psychology, Fordham University

CHARLES E. SCHAEFER

A. INTRODUCTION

In recent years the Gough Adjective Check List (ACL) has been employed with increasing frequency in psychological research. The main value of this instrument for personality assessment is that it offers a unique picture of a person's self-concept by presenting an extensive checklist of adjectives commonly used in everyday life: for example, spunky, good-looking, smug. The motivational significance of the self-concept in determining a person's behavior has been pointed out by Rank (16).

Many of the currently available scales for the ACL seem particularly relevant for research on the creative person. Indeed, studies from the Institute of Personality Assessment and Research (12) have indicated that creative adults tend to score high on such ACL scales as Autonomy and Aggression, while scoring comparatively low on such scales as Abasement, Affiliation, Deference. In terms of individual adjectives, MacKinnon (11) reports that 80 percent or more of a group of creative architects checked the following adjectives as self-descriptive: imaginative (98 percent); active (92 percent); honest, idealistic, and inventive (90 percent). MacKinnon and his associates further report that staff observations verified the accuracy of the ACL self-descriptions.

The purpose of the present study was to extend creativity research with the ACL to a younger population: namely, high school students. Several authors (20, 21, 22) have pointed to the scarcity of creativity investigations with this age group. Yet, as Hammer observed, "Clinical experience suggests that psychological techniques administered to adolescent subjects are frequently more revealing of the underlying traits and feelings than are tests administered to adults" (8, p. 327).

---

* Received in the Editorial Office, Provincetown, Massachusetts, on May 15, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This study is part of a larger project supported by Subcontract No. 2 of the Center for Urban Education's Contract OEC-1-7-062868-3060 with the United States Office of Education.

B. METHOD

1. *Subjects*

The subjects were 800 male and female students from 10 high schools in the New York metropolitan area.[2] The schools were chosen because they offer courses or programs providing opportunities for creative achievement, such as creative writing or science projects. The subjects included 434 seniors, 272 juniors, and 94 sophomores. The group as a whole was superior in terms of academic achievement and intelligence test scores.

2. *Creativity Criteria*

The subjects were assigned to eight criterion groups of 100 students each, according to the following threefold system of classification: achievement (creative or control), field of achievement, and sex. For the boys, the achievement fields were broadly classified as either art-writing or scientific. The art-writing field for the boys included all forms of visual arts and all types of literary expression, while the scientific field encompassed the natural sciences and mathematics. Since few girls were found who produced a scientific work, the girls' specialty fields were classified separately as either art or writing. The eight criterion groups were designated: Creative-Art-Writing-Boys (CrAW-B), Control-Art-Writing-Boys (CoAW-B), Creative-Scientific-Boys (CrS-B), Control-Scientific-Boys (CoS-B), Creative-Art-Girls (CrA-G), Control-Art-Girls (CoA-G), Creative-Writing-Girls (CrW-G), Control-Writing-Girls (CoW-G).

The criterion employed for assessing creativity was a combination of teachers' evaluations and creativity test scores. Accordingly, the creative *S*s were nominated by their teachers as having produced one or more creative products, such as drawings, poems, stories, science projects. They also scored above a cutoff point on Guilford's Alternate Uses and Consequences tests. The control *S*s were nominated by the same teachers as having shown no concrete evidence of creativity, and they scored below a cutoff point on the Guilford screening tests. Within each of the four specialty fields, the creative and control *S*s were matched in school attended, class in which enrolled, grade level, and grade-point average.

3. *Procedure*

The ACL was administered as part of a larger test battery during a two-hour testing session held outside of school hours. The *S*s were paid for participation in this study. Identification numbers were employed during the testing session to provide anonymity, and the students were assured of the confidentiality of their responses. The ACL was scored for each of the 24 currently available experimental scales.

---

[2] The participating high schools were Abraham Lincoln, Art and Design, Bronxville, Erasmus Hall, Forest Hills, Jamaica, Midwood, Music and Art, Regis, and Stuyvesant.

TABLE 1
MEANS AND STANDARD DEVIATIONS OF BOYS' CRITERION GROUPS ON ADJECTIVE CHECK LIST

	Criterion groups							
	Art-Writing				Scientific			
Adjective Check List scales	Creative		Control		Creative		Control	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Creative groups hypothesized higher								
No. Unfavorable Adj.	55.48	9.57	54.27	11.33	55.24	11.58	53.04	10.91
Self-confidence	49.45	10.00	47.24	9.88	51.41	9.17	46.30	9.92
Lability	57.51	8.96	50.48	10.00	55.57	12.10	50.73	8.41
Dominance	50.37	11.42	49.14	10.53	52.23	10.25	49.22	10.14
Exhibition	55.64	9.63	51.29	11.42	55.39	10.24	50.66	11.07
Autonomy	56.80	9.47	49.82	9.90	54.98	9.71	49.16	8.60
Aggression	56.29	9.55	52.34	10.14	57.08	11.15	52.47	9.64
Change	55.59	10.21	47.74	10.06	51.19	10.38	48.31	9.67
Control groups hypothesized higher								
Defensiveness	44.68	8.92	47.37	7.86	46.78	9.77	47.32	8.84
Self-control	40.62	9.82	45.72	10.17	43.40	11.09	46.12	9.99
Order	44.84	11.21	49.70	9.77	49.42	11.77	49.15	9.67
Nurturance	45.13	10.69	47.78	10.06	42.56	11.67	47.27	10.05
Succorance	48.78	9.90	50.94	10.77	49.74	10.35	51.10	10.02
Deference	42.64	10.27	48.96	11.09	42.55	10.19	47.93	9.58
No specific hypotheses								
Number checked	48.24	8.09	47.12	8.98	48.65	7.77	47.32	8.90
No. Favorable Adj.	43.90	9.82	45.05	9.52	44.52	10.65	44.39	9.97
Personal Adjustment	41.25	9.44	44.18	9.78	43.13	10.58	45.26	9.11
Achievement	49.29	11.33	49.37	9.69	52.71	10.92	50.69	9.45
Intraception	49.66	10.32	48.79	10.65	49.24	11.39	48.10	9.28
Endurance	46.52	10.92	50.71	9.40	50.27	12.17	51.14	10.23
Heterosexuality	47.81	11.75	49.19	10.35	44.55	10.69	46.71	10.23
Affiliation	45.25	9.48	46.99	9.07	43.94	10.25	45.59	9.27
Abasement	48.29	10.46	50.62	9.41	47.46	9.57	50.86	9.71
Counseling Readiness	49.19	10.20	50.57	9.56	49.09	9.89	51.66	10.83

C. RESULTS

The means and standard deviations of the criterion groups on the ACL are presented in Tables 1 and 2. On the basis of a review of the literature, it was hypothesized that the creative students would score significantly higher than the controls on the following scales: Number of unfavorable adjectives, Self-confidence, Lability, Dominance, Exhibition, Autonomy, Aggression, and Change. It was further hypothesized that the control *S*s would score higher on the scales of Defensiveness, Self-control, Order, Nurturance, Succorance, and Deference. The ACL group comparisons found in Table 3 indicate that the mean differences on the above 14 scales were almost invariably in the hypothesized direction. The direction of mean differences for the ten remaining ACL scales was not consistent.

TABLE 2
MEANS AND STANDARD DEVIATIONS OF GIRLS' CRITERION GROUPS ON ADJECTIVE CHECK LIST

	Criterion groups							
	Art				Writing			
Adjective Check List scales	Creative		Control		Creative		Control	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Creative groups hypothesized higher								
No. Unfavorable Adj.	57.68	13.04	54.71	11.90	56.75	13.04	51.92	10.48
Self-confidence	48.35	8.91	46.58	9.63	50.73	10.11	47.89	10.12
Lability	54.03	10.07	51.29	9.25	55.46	10.71	48.74	10.06
Dominance	45.88	10.20	45.27	10.85	47.25	10.59	48.03	9.79
Exhibition	52.47	10.99	50.99	10.15	54.41	10.62	50.30	9.84
Autonomy	54.16	10.50	52.62	10.53	57.10	10.74	49.10	8.64
Aggression	55.58	10.59	53.94	10.80	57.80	10.25	52.37	9.52
Change	54.18	10.13	51.21	10.05	58.05	11.04	50.04	10.24
Control groups hypothesized higher								
Defensiveness	40.70	12.11	43.30	11.44	39.48	12.45	47.08	9.75
Self-control	43.61	10.31	45.73	9.70	42.02	11.21	49.76	9.45
Order	44.16	11.04	45.76	8.54	42.20	11.60	49.61	9.25
Nurturance	42.72	12.60	44.64	11.77	42.18	12.52	48.62	10.27
Succorance	53.46	10.54	53.79	10.85	54.43	11.13	53.20	8.82
Deference	44.24	10.42	46.33	9.87	41.37	10.37	48.92	9.74
No specific hypotheses								
Number checked	55.48	10.50	50.96	9.65	56.50	12.27	50.54	9.08
No. Favorable Adj.	47.17	11.20	47.07	12.11	46.39	11.26	49.87	10.24
Personal Adjustment	42.38	10.55	43.92	11.36	41.82	10.26	47.26	9.70
Achievement	45.77	9.53	45.35	9.92	47.23	10.25	49.15	9.43
Intraception	46.47	9.38	47.26	10.95	47.72	8.40	50.05	9.14
Endurance	44.51	10.64	45.96	8.78	43.53	11.35	49.93	8.94
Heterosexuality	49.03	11.89	48.28	10.99	47.80	11.61	49.83	9.92
Affiliation	40.79	11.70	41.69	11.14	40.88	12.24	45.03	10.05
Abasement	51.39	10.93	52.56	11.01	50.51	10.20	52.36	9.14
Counseling Readiness	55.32	10.96	54.31	11.19	57.98	10.10	53.18	10.51

In view of the observation by Gough and Heilbrun that "in many situations the greatest value of the ACL may accrue from noting, pondering, or analyzing those specific words which an individual or group of individuals has checked as self-descriptive" (7, p. 4), an item analysis was conducted to identify the individual adjectives that differentiated the creative and control *S*s in each specialty field. Following the item analysis procedure outlined in Schaefer and Anastasi (19), a compound probability value was determined

TABLE 3

NORMAL CURVE RATIOS[a] OF MEAN CRITERION GROUP DIFFERENCES ON ADJECTIVE CHECK LIST

	Criterion groups[b]			
Scales	CrAW-B *vs.* CoAW-B	CrS-B *vs.* CoS-B	CrA-G *vs.* CoA-G	CrW-G *vs.* CoW-G
Creative groups hypothesized higher[c]				
No. Unfavorable Adj.	.82	1.38	1.68*	2.89**
Self-confidence	1.57	3.78***	1.35	1.98*
Lability	5.24***	3.29***	2.00*	4.57***
Dominance	.79	2.09*	.41	.54
Exhibition	2.92**	3.12***	.99	2.84**
Autonomy	5.09***	4.48***	1.03	5.80***
Aggression	2.84**	3.14***	1.08	3.88***
Change	5.49***	2.03*	2.08*	5.32***
Control groups hypothesized higher[c]				
Defensiveness	−2.26*	− .41	−1.56	−4.81***
Self-control	−3.62***	−1.82*	−1.50	−5.28***
Order	−3.26***	.18	−1.15	−4.99***
Nurturance	−1.80*	−3.06**	−1.11	−3.98***
Succorance	−1.48	− .94	− .22	.86
Deference	−4.18***	−3.84***	−1.45	−5.31***
No specific hypotheses				
Number checked	.92	1.13	3.17**	3.90***
No. Favorable Adj.	− .84	.09	.06	−2.28*
Personal Adjustment	−2.15*	−1.52	− .99	−3.85***
Achievement	− .05	1.40	.30	−1.38
Endurance	−2.91**	− .55	−1.05	−4.43***
Intraception	.59	.78	− .55	−1.88
Affiliation	−1.33	−1.20	− .56	−2.62**
Heterosexuality	− .88	−1.76	.46	−1.33
Abasement	−1.65	−2.50*	− .75	−1.35
Counseling Readiness	− .98	−1.75	.64	3.29**

[a] Negative signs indicate higher mean scores for control group.

[b] CrAW-B = Creative-Art-Writing-Boys, CoAW-B = Control-Art-Writing-Boys, CrS-B = Creative-Scientific-Boys, CoS-B = Control-Scientific-Boys, CrA-G = Creative-Art-Girls, CoA-G = Control-Art-Girls, CrW-G = Creative-Writing-Girls, CoW-G = Control-Writing-Girls.

[c] One-tailed tests of significance used.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

for each of the 300 adjectives of the ACL. Those adjectives which significantly differentiated the creative and control groups at the compound probability level of .10 or better in one or more of the four specialty fields were identified.

The results of this item analysis can be summarized as follows. The creative adolescents, significantly more often than the controls, viewed themselves as creative (imaginative, artistic, ingenious, original), independent (unconventional, individualistic), uninhibited (spontaneous, impulsive), iconoclastic (assertive, outspoken, rebellious), complicated (complex, reflective, cynical, idealistic), and asocial (aloof). In striking contrast, the adjectives differentiating the control students present a remarkably consistent picture of passive conformity: dependable, cooperative, contented, conventional, quiet, and silent.

D. Discussion

The results of the present study substantiate the impressive record of the ACL in creativity research. The high level of validity found with a number of ACL scales in previous investigations of creative adults (4, 13) was confirmed in this study of creative adolescents. The ability of a number of ACL scales to identify creative adolescents of both sexes and across diverse fields of achievement indicates that the scales have considerable validity generalization. The results support the observation of Cashdan and Welsh (3) that creative adolescents are very similar in personality despite sex and specialty differences. This similarity underscores the feasibility of Golann's (6) suggestion that personality correlates be used as criterion measures in the study of creativity.

1. *The Self-Image of Creative Adolescents*

Since recent factor analyses of the ACL scales (9, 15, 18) have pointed to a substantial degree of overlap in the 24 scales, the results of the present study can be described in terms of general factors or themes. The ACL scales found to characterize the creative *S*s in this study suggest the following main themes: (*a*) complexity and reconciliation of opposites; (*b*) impulsivity and craving for novelty; and (*c*) autonomy and self-assertion.

a. Complexity and reconciliation of opposites. The openness of the creative adolescent to many sides of his personality, including socially undesirable traits, is apparent in the finding that the creative *S*s scored high on the ACL scales of Number of Adjectives Checked and Number of Unfavorable Adjectives, while they scored low on Defensiveness and Personal Adjustment.

Consonant with this theme are a number of the individual adjectives found to be distinctive of the creative *S*s: namely, complex, assertive, aloof, cynical, idealistic, impulsive, reflective, and rebellious. These adjectives suggest that the creative person not only recognizes opposing forces in his nature, such as cynical *vs.* idealistic, impulsive *vs.* reflective, and aloof *vs.* assertive, but is able to reconcile and integrate them into his personality. To do this the creative *S*s must possess exceptionally strong egos. In this connection, Barron (2, p. 159) has described the creative person as "both more primitive and more cultured, more destructive and more constructive, crazier and saner, than the average person." The control *S*s, on the other hand, were a picture of calm contentment, as reflected in the adjectives that characterized them, such as cooperative, contented, and quiet.

The number of unfavorable adjectives found to differentiate the creative and the control groups raises the question as to whether the creative *S*s were simply more open and honest in their responses: that is, more open to self-disclosure behavior as described by Jourard (10), or whether the adjectives reflected real personality differences between the *S*s in the two groups. The discrepancy between the calm, quiet, and contented self-portrait of the control *S*s and the typical depiction of the adolescent years as stormy and stressful would support the former view. On the other hand, MacKinnon's (11) finding that staff observations tend to verify ACL self-descriptions would support the latter theory. Further research is needed to resolve this question.

b. Impulsivity and craving for novelty. Closely related to the complex, conflictful self-concept of the creative *S*s is the second general theme: namely, openness to impulse expression and desire for novel experiences. This theme is reflected in the high scores of the creative students on the scales of Lability, Exhibition, and Change, and their low scores on Endurance, Order, and Self-control. The following individual adjectives found to be distinctive of the creative *S*s are also indicative of this theme: impulsive, imaginative, spontaneous, artistic, ingenious, and original. In brief, the creative *S*s see themselves as continually striving for novelty in their environment, both by impulse expression and more deliberate creative works. Rees and Goldman (17) also found that creativity is more apt to occur in college students who are impulsive and characterized by a lack of deliberation, self-control, and restraint. They speculate that the impulsive, acting out, histrionic behavior of the creative person is indicative of a basic hysterical character structure. A motivational relationship between openness to novelty and creative achievement has been suggested by Maddi (14).

c. Autonomy and Self-Assertion. The third general component of the self-concept of the creative adolescents is their independence of thought and action, coupled with a leadership role in social situations. Thus, the creative *S*s scored high on the ACL scales of Autonomy, Aggression, Dominance, and Self-confidence, while they scored relatively low on Abasement, Affiliation, Deference, and Nurturance. The results of the item analysis showed the creative *S*s to be distinguishable by their self-image as aloof, individualistic, and unconventional. The creative adolescent, then, prefers to work by himself, although he tends to be poised, influential, and controlling in social situations.

In view of the above, it is not surprising that an analysis of the *S*s' biographical data revealed that more creative than control *S*s were elected to the highest position of leadership in high school extracurricular activities (1, 19). Similarly, the results of a factor analytic study by Furcon, Baehr, and Zolik (5) disclosed that creative persons tend to relate to their social milieu by dominating and communicating with groups, yet otherwise remaining socially aloof and demonstrating independence of thought and action. It has been often found that this independence is carried over into the work environment by a preference for highly unstructured work situations which foster autonomy, freedom, and an idiosyncratic approach to work activity (5, 12).

2. *Sex and Specialty Differences*

Although the four creative groups in this study tended to show similar patterns of scores on the ACL scales, there were obvious sex and specialty differences in scale scores between the four groups. The most conspicuous variation was that relatively few significant differences between creative and control group means were found in the girls' art field. This finding is undoubtedly related to the fact that the majority of creative and control girls in this field attended high schools that specialized in art. Thus, the girls were already a highly select group in regard to creativity so that the creative-control distinction in this field was rather tenuous.

An obvious sex difference in ACL scores was that the creative girls (as compared with their controls) tended to check more adjectives as self-descriptive, particularly unfavorable ones, than the creative boys (as compared with their controls). This difference was particularly noticeable for the girls in the writing field. Perhaps the simplest explanation for this finding is that girls tend to be more verbal than boys and one would expect girls with writing talent to be especially verbal. Also relevant is Trezise's (23) finding that creative girls tend to be more introspective and interested in self-understanding than creative boys.

E. SUMMARY

The Adjective Check List was found to be a highly valid instrument in identifying creative high school boys and girls in different specialty fields. As contrasted with matched control groups, creative adolescents were found to have similar self-images which were described in terms of three themes: complexity and reconciliation of opposites, impulsivity and craving for novelty, and autonomy and self-assertion. It was concluded that a person's self-image plays a significant role in motivating him towards creative achievement.

REFERENCES

1. ANASTASI, A., & SCHAEFER, C. E. Biographical correlates of artistic and literary creativity in adolescent girls. *J. Appl. Psychol.*, 1969, **57**, in press.
2. BARRON, F. The needs for order and for disorder as motives in creative activity. In C. W. Taylor & F. Barron (Eds.), *Scientific Creativity: Its Recognition and Development.* New York: Wiley, 1963.
3. CASHDAN, S., & WELSH, G. S. Personality correlates of creative potential in talented high school students. *J. Personal.*, 1966, **34**, 445-455.
4. CRUTCHFIELD, R. S. Independent thought in a conformist world. In S. M. Farber & R. H. C. Wilson (Eds.), *Conflict and Creativity.* New York: McGraw-Hill, 1963. Pp. 208-237.
5. FURCON, J. E., BAEHR, M. E., & ZOLIK, E. S. Creative personality: A factor analytic study. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, Illinois, May 1966.
6. GOLANN, S. E. Psychological study of creativity. *Psychol. Bull.*, 1963, **60**, 548-565.
7. GOUGH, H. G., & HEILBRUN, A. B. The Adjective Check List manual. Palo Alto, Calif.: Consulting Psychologists Press, 1965.
8. HAMMER, E. F. Personality patterns in young creative artists. *Adolescence*, 1967, **1**, 327-350.
9. HANDY, D. C. P. The relationship of hierarchical need level to success of Peace Corps trainees. Unpublished Doctoral dissertation, University of Texas, Austin, Texas, 1966.
10. JOURARD, S. M. A study of self-disclosure. *Sci. Amer.*, 1958, **198**, 77-82.
11. MACKINNON, D. W. Creativity in architects. In *The Creative Person.* Berkeley, Calif.: Univ. Calif. Press, 1961. Ch. 5.
12. ———. Personality and the realization of creative potential. *Amer. Psychol.*, 1965, **20**, 273-281.
13. ———. Some reflections on the current status of personality assessment with special reference to the assessment of creative persons. Paper presented at the Faculty Symposium, Department of Psychology, University of California, Berkeley, California, November 1966.
14. MADDI, S. R. Motivational aspects of creativity. *J. Personal.*, 1965, **33**, 330-347.
15. PARKER, G. V. C., & MEGARGEE, E. I. Factor analytic studies of the Adjective Check List. *Proc. 75th Ann. Convention Amer. Psychol. Assoc.*, 1967, **2**, 211-212.
16. RANK, O. Will therapy and truth and reality. (Trans. by J. Taft.) New York: Knopf, 1945.
17. REES, M. E., & GOLDMAN, M. Some relationships between creativity and personality. *J. Gen. Psychol.*, 1961, **65**, 145-161.

18. Scarr, S. The Adjective Check List as a personality assessment technique with children. *J. Consult. Psychol.*, 1966, **30**, 122-128.

19. Schaefer, C. E., & Anastasi, A. A biographical inventory for identifying creativity in adolescent boys. *J. Appl. Psychol.*, 1968, **52**, 42-48.

20. Smith, R. M. Characteristics of creativity research. *Percept. & Motor Skills*, 1968, **26**, 698.

21. Taylor, C. W. (Ed.). Creativity: Progress and Potential. New York: McGraw-Hill, 1964.

22. Torrance, E. P. Guiding Creative Talent. Englewood Cliffs, N.J.: Prentice-Hall, 1962.

23. Trezise, R. L. A descriptive study of the life styles of a group of creative adolescents. *Diss. Abst.*, 1967, **27** (9-A), 2754-2755.

Department of Psychology
Fordham University
Bronx, New York 10458

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 243-246.

DEVELOPMENTAL STUDY OF SOCIAL ROLE PERCEPTION AMONG HUTTERITE ADOLESCENTS*[1]

Department of Psychology, University of Manitoba

Shirin Schludermann and Eduard Schludermann

A. Introduction

In a previous study of the social role perception among Hutterite children, Schludermann and Schludermann (2) suggested, on the basis of results obtained, that as Hutterite children grow older, they may be able to make the kind of differentiations in their perceptions that researchers using the semantic differential are accustomed to obtaining. The present study on Hutterite adolescents was done with a view to testing the initial assumption that there are a few basic dimensions underlying Hutterite adolescents' perception of social roles in their communal society. The second part of the assumption was that the basic dimensions would emerge as clearly differentiated dimensions which would indicate a certain degree of change in the developmental differentiation in cognition. It is generally assumed that an older person is more capable of making finer differentiation in perceptual fields than a younger person. Werner (3) and Piaget [see Flavell (1)] have demonstrated the validity of such a developmental change in cognitive differentiation.

B. Method

1. *Subjects*

One hundred and twenty-eight adolescents were selected from the public schools in 20 Hutterite colonies of Manitoba. The subjects were 14 to 16 years old and consisted of only those adolescents who were average or above-average in their abilities. The selection was made on the basis of their school achievement records and teachers' evaluation. Extremely dull and outstanding subjects were eliminated in the selection so that the effects of extreme differences in intelligence could be somewhat controlled. The Hutterite colonies

---

* Received in the Editorial Office, Provincetown, Massachusetts, on May 16, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This study was supported by a grant from the Laidlaw Foundation.

are located in rural areas and form part of the socioeconomic homogeneous Hutterite communal society.

2. *The Instrument, Concepts, and Scales*

The description of the semantic differential scales, concepts, and their method of selection were kept the same as in the previously reported study (2). Therefore, they are not repeated here. The investigative procedure was also the same as in the previous study.

C. RESULTS

The sample was subdivided into two groups of girls ($N = 76$) and boys ($N = 52$). Correlation matrices of 15 × 15 scales for the two subgroups were calculated separately and factor analyzed by the principal axes method.

The principal axes matrices obtained for the two groups are similar to those obtained for Hutterite children. Table 1 shows that only the first factor in both groups accounts for one-third of the total variance. The second factor accounts for 12 to 13 percent and the third factor accounts for seven percent variance. Thus the first three factors account for approximately only 50 percent of the total variance. It is clear that with the exception of the first factor, the rest of the factors are specific factor variances. The first factor, which qualifies as a common factor variance, appears to be an undifferentiated global factor. Twelve of the 15 variables have loadings higher than ±.40 on the first factor of the girls' matrix, and ten of the 15 variables have similar loadings on the first factor of the boys' matrix. Thus the dominant dimension appears to be essentially "good-bad," and any person who is Interesting, Brave, Busy, Bright, Useful, Pleasant, etc. is a "good" person. The second factor consists of four variables: important, friendly, clean, and active; whereas the third factor has a high loading only on one variable in both groups. The reader will be impressed with the similarity of factor matrices obtained across age groups of Hutterite children eight years to 13 years (in previous studies), and adolescents 14 to 16 years in the present study.

TABLE 1
PROPORTION OF VARIANCE ACCOUNTED FOR BEFORE ROTATION

Sex	Ages	F_1	F_2	F_3	F_4	Cumulative percent variance
Girls	14-16 yrs.	30.71	11.96	7.09	6.37	56.13
Boys	14-16 yrs.	32.25	13.69	7.12	6.48	59.54

D. DISCUSSION

1. *Factor Structure*

It was suggested previously (2) that such low common factor variances may be due to several reasons. One is that the subjects did not understand the instructions clearly; and a second is that the subjects had only a vague understanding of the descriptive terms. These two sources of error may be expected to be less influential in the case of the adolescents than in the case of the children. The task itself was kept simple enough for a child to follow, and the descriptive adjectives were carefully selected so that a child as young as eight years old would be able to understand the meaning of the words. Therefore, it may be assumed that these two sources of error which might have reduced the common factor variances in the case of the children, would not necessarily have had such drastic effects in the case of the adolescents. A third source of error variance was suggested, and that is the peculiarities of the social environment in which Hutterite children and adolescents grow up. Hutterite adolescents' social interactions are strictly supervised and limited to visits with members of a few other colonies. It is therefore quite likely that the Hutterite adolescents, when asked to describe certain social roles in their colonies, may tend to think more in terms of the two or three other farm-foremen or pigmen they know. They may further tend to think of these individuals in terms of their personal characteristics rather than in terms of social function and role. It is very likely, therefore, that the peculiarities of social environment tend to exert a very strong influence on the direction and degree to which Hutterite adolescents will be able to direct their thinking and perception of certain relevant or irrelevant characteristics of their environment.

2. *Social Perception*

In terms of the number of common factors that may be used as indices of cognitive differentiation (i.e., the greater the number of common factors, the greater the cognitive differentiation), Hutterite adolescents fail to show any clear differentiation of social roles. The Hutterite adolescents' undifferentiated social perception is not much different from that of the Hutterite children. It may be asked as to why the Hutterite adolescents' cognitive growth fails to show the expected concomitant changes towards cognitive differentiation. Here again, it is important to remember that Hutterite children and adolescents have to live in a society which has deliberately abolished all signs of status symbol and which offers very little positive reinforcement for making distinctions between social roles, such as superior-inferior, etc. Very often, there is considerable negative reinforcement among the Hutterites against attempts to make such distinctions outwardly, and to publicize openly through dress, residence, and mode of behavior, one's superior status. It appears then that in Hutterite society failure to obtain developmental differentiation in the social role perception of Hutterite adolescents is very likely due to the powerful social influences that shape the growing personality of the young ones. If this assumption is true, then one may expect to find similarly undifferentiated perception of social roles even among the Hutterite adults.

E. SUMMARY

One hundred and twenty-eight Hutterite adolescents, 14 to 16 years of age, were selected from 20 colonies in Manitoba. The adolescents described the same 16 social roles on 15 semantic scales as the Hutterite children did. The two correlation matrices of 15 × 15 for two sex groups were factor analyzed separately. There was only one common factor which accounted for one-third of the total variance. The principal factor matrix of the adolescents was almost as undifferentiated as that of the children. The assumption of a few basic dimensions underlying adolescents' perception, as well as the assumption of increased cognitive differentiation in social role perception of adolescents, were not supported.

REFERENCES

1. FLAVELL, J. The Developmental Psychology of Jean Piaget. Princeton, N.J.: Van Nostrand, 1963.
2. SCHLUDERMANN, S., & SCHLUDERMANN, E. Social role perceptions of children in Hutterite communal society. *J. of Psychol.*, 1969, **72**, 183-188.
3. WERNER, H. Comparative Psychology of Mental Development. New York: Follet, 1948.

Department of Psychology
University of Manitoba
Winnipeg, Canada

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 247-251.

RAVEN'S PROGRESSIVE MATRICES: MORE ON CONSTRUCT VALIDITY*

East Orange Veterans Administration Hospital and Rutgers—The State University

Henry R. Burke and William C. Bingham

A. Introduction

A previous study (2) cited evidence that Raven's Progressive Matrices had validity as a measure of general intellectual functioning. It loaded significantly on a factor of general intelligence in an analysis of the intercorrelations of scores on the Progressive Matrices, on Cohen's factors (3, 4), on the WAIS (7), and on the separate parts of the Army General Classification Test (1). Because of the small sample size in that study ($N = 39$) interpretations had to be highly tentative.

B. Procedures

For the present study, the senior author himself accumulated 52 additional protocols, using the same tests. The Progressive Matrices was administered *without a time limit*. The other tests were administered in standard fashion. The WAIS was scored also for Cohen's factors (Verbal Comprehension, Memory, and Perceptual Organization). The AGCT was scored separately for Vocabulary, Arithmetic, and Block Counting.

All subjects were male patients at the East Orange Veterans Administration Hospital who were referred to the Psychology Service for vocational counseling. Ages ranged from 19 to 59 with a mean of 35.1 and *SD* 8.7. The 91 sets of test scores were intercorrelated and subjected to a principal components analysis and a varimax rotation on electronic data processing equipment at the Center for Computer and Information Services, Rutgers University.

C. Findings and Discussion

1. *Reliability of the Progressive Matrices*

Perhaps the most surprising finding in this study was the correlation between the 91 pairs of odd-even scores on the Progressive Matrices: $r = .92$. Application of the Spearman-Brown formula for predicting the reliability of a test of double length (5, p. 194) yielded a reliability coefficient of .96.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on May 20, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

TABLE 1
INTERCORRELATIONS OF TEST SCORE VARIABLES
($N = 91$)

Variables	1	2	3	4	5
1. WAIS Verbal*					
2. WAIS Perf.*	70				
3. WAIS Total*	95	90			
4. AGCT Total**	80	71	82		
5. Raven PM**	65	76	75	67	
M	60.9	45.3	106.1	104.0	41.2
SD	14.2	10.4	22.8	19.6	11.5

Note: Decimals omitted.
* Scaled scores.
** Raw scores.

2. *Concurrent Validity*

In this study, the Progressive Matrices, as an apparently nonverbal and untimed measure of "intellectual capacity" (6, p. 1), correlated slightly better (Table 1) with the WAIS Performance ($r = .76$) than did either the WAIS Verbal ($r = .70$) or the AGCT Total ($r = .71$). The Raven PM did not correlate quite as well with the WAIS Total ($r = .75$) as did the AGCT Total ($r = .82$). This fact seems to be attributable to the highly verbal character of both the WAIS and the AGCT: WAIS-Verbal and AGCT Total, $r = .80$. As further evidence, Table 2 gives a paired comparison of the correlations of the Raven PM and AGCT scores with the WAIS subtest scores. However, the Raven PM correlated about as well with the WAIS Verbal ($r = .65$) and the WAIS Performance ($r = .76$) as these two major parts of the WAIS did with each other ($r = .70$), as seen in Table 1.

TABLE 2
INTERCORRELATIONS OF WAIS SUBTEST SCALED SCORES WITH
RAVEN PM AND AGCT TOTAL SCORES
($N = 91$)

WAIS subtests	Raven PM	AGCT
Information	56	69
Comprehension	50	61
Arithmetic	43	61
Similarities	59	69
Digit Span	47	54
Vocabulary	55	70
Digit Symbol	51	52
Picture Completion	64	59
Block Design	60	56
Picture Arrangement	50	43
Object Assembly	62	61

Note: Decimals omitted.

TABLE 3
INTERCORRELATIONS OF TEST SCORE VARIABLES AND AGE
($N = 91$)

Variables	1	2	3	4	5	6	7	8
1. Age								
2. Cohen V	03							
3. Cohen M	—04	63						
4. Cohen PO	—11	51	52					
5. AGCT V	06	77	60	56				
6. AGCT N	—14	72	69	56	81			
7. AGCT BL	—18	45	45	59	58	68		
8. Raven PM	—32	59	49	61	60	66	55	
M	35.1	10.3	9.8	9.8	26.7	28.2	25.5	41.3
SD	8.7	2.8	2.7	2.7	9.0	9.1	8.8	11.5

Note: Decimals omitted.

The intercorrelations in this study ($N = 91$) of the Raven PM with the WAIS and AGCT, and of the WAIS with the AGCT, are somewhat less remarkable than the same correlations ($N = 39$) in the previous study (2): Raven PM and WAIS, $r = .75$ *vs.* .85; Raven PM and AGCT, $r = .67$ *vs.* .70; WAIS and AGCT, $r = .82$ *vs.* .84.

3. *Construct Validity*

There appeared to be too many variables affecting the intercorrelations between the Raven PM, the WAIS Verbal, the WAIS Performance, and the AGCT (Table 1) to explain their relationships in terms of a *single* general factor of intelligence, if one applied the Spearman criterion of zero tetrad differences (5, pp. 248 ff.).

However, when the intercorrelations (Table 3) among the Raven PM, Cohen's WAIS factors, the AGCT part scores, and age were submitted to a principal components analysis, all variables except age loaded heavily (Table 4) on a single factor of general intellectual functioning. It appears that all

TABLE 4
PRINCIPAL COMPONENTS ANALYSIS OF VARIABLES
($N = 91$)

Variables	A	B	C	D	E
Age	—13	53	19	—05	—08
Cohen V	80	29	—15	12	—09
Cohen M	74	15	—19	—31	06
Cohen PO	72	—16	13	—13	—26
AGCT V	86	26	11	15	03
AGCT N	90	02	—02	02	22
AGCT BL	72	—25	31	—04	10
Raven PM	76	—29	—14	12	—11

of these tests, including the Raven PM, are somehow, in spite of their differences, all measures of general intelligence.

The same matrix of intercorrelations (Table 3) was submitted to varimax rotation, and five factors were identified (Table 5). Since the sample size was relatively small ($N = 91$), it seemed appropriate to employ a rigorous criterion for loading. On that basis a cut-off of .40 was employed in determining factor content.

On Factor A, the only variables which did not load were age (.03), Cohen's Perceptual Organization (.34), and AGCT Block Counting (.31). Although age clearly did not contribute, the other two barely missed meeting the criterion. Factor A also appeared to be a general factor of intellectual functioning, predominantly verbal.

TABLE 5
ROTATED FACTOR MATRIX
(Number of Iteration Cycles 6)

			Factors		
Variables	A	B	C	D	E
1. Age	03	—59	—07	—01	—04
2. Cohen V	79	—03	16	27	24
3. Cohen M	47	05	22	62	19
4. Cohen PO	34	15	41	23	53
5. AGCT V	76	—09	43	19	18
6. AGCT N	66	18	49	36	08
7. AGCT BL	31	21	68	16	24
8. Raven PM	53	44	28	16	36

Note: Decimals omitted. $N = 91$.

Factor B loaded only on age (—.59) and the Raven PM (.44), and appeared to be an age factor. This finding supported that of our earlier study (2), and Raven's contention of a negative relationship between scores on his test and age.

Factor C loaded on Cohen's Perceptual Organization and all of the AGCT parts. Since these four tests were all timed, and the AGCT parts highly so, this factor seemed to be clearly a time or speed factor.

Factors D and E appeared to be similar to Cohen's Memory and Perceptual Organization Factors, respectively, since those variables were the only ones which loaded on these two factors. It is interesting to note that the Raven PM missed loading (.36) on Factor E (Perceptual Organization), and, accordingly, as in our previous study (2), did not appear to be uniquely or clearly significantly a measure of this Cohen factor, as we hypothesized it might be. Possibly Factor D (Cohen's Memory Factor) might need to be thought of as a Number factor.

Although Factor A, which we regard as a general factor of intellectual functioning, was highly verbal, the Raven Progressive Matrices loaded on it. One might speculate as to what extent performance on the Raven PM depended on verbal ability. The investigators have noted that some subjects, especially on the more difficult items, appeared to "talk their way" through the Progressive Matrices: that is, they seemed to explain to themselves verbally what the relationships were. It might be interesting to investigate this kind of phenomenon more systematically.

A varimax rotation, in summary, showed that the Raven Progressive Matrices was positively related to a general factor of intellectual functioning, even though it was highly verbal, negatively related to an age factor, unrelated to speed and memory factors, and no more than minimally related to a perceptual organization factor.

D. Summary

(*a*) The untimed Raven Progressive Matrices showed a surprising split-half corrected reliability coefficient of .96 ($N = 91$). (*b*) Both a principal components and a varimax factor analysis indicated that the Raven Progressive Matrices loaded significantly on a factor of general intellectual functioning, even when it was heavily determined by verbal content. (*c*) Performance on the Raven Progressive Matrices was negatively related to age.

References

1. Army General Classification Test, Technical Report. Chicago, Ill.: Sci. Res. Assoc. 1960.
2. Bingham, W. C., Burke, H. R., & Murray, S. Raven's Progressive Matrices: Construct validity. *J. of Psychol.*, 1966, **62**, 205-209.
3. Cohen, J. A factor-analytically based rationale for the Wechsler Adult Intelligence Scale. *J. Consult. Psychol.*, 1957, **21**, 451-457.
4. ———. The factorial structure of the WAIS between early adulthood and old age. *J. Consult. Psychol.*, 1957, **21**, 283-290.
5. Peters, C. C., & Van Voorhis, W. R. Statistical Procedures and Their Mathematical Bases. New York: McGraw-Hill, 1940.
6. Raven, J. C. Guide to Using Progressive Matrices (1938). London, England: H. E. Lewis, 1956.
7. Wechsler, D. Manual for the Wechsler Adult Intelligence Scale. New York: Psychol. Corp., 1955.

197 Ridgewood Avenue
Glen Ridge, New Jersey 07028

Rutgers—The State University
New Brunswick, New Jersey 08903

Published as a separate and in *The Journal of Psychology*, 1969, 72, 253-261.

SCALE CHECKING STYLE AS A FUNCTION OF AGE AND SEX IN INDIAN AND HUTTERITE CHILDREN*[1]

Department of Psychology, University of Manitoba

SHIRIN SCHLUDERMANN AND EDUARD SCHLUDERMANN

A. INTRODUCTION

Werner (16) emphasized that a young organism in the process of development goes through increased differentiation, hierarchization, and integration. Werner had proposed the principle of differentiation with special reference to cognitive and mental development. Developmental psychologists have endeavored to enunciate and define a variety of criteria whereby differentiation in children and adults may be studied. The development of cognitive differentiation in young children is one such area where a considerable number of criteria and hypotheses are generated and tested (e.g., 2, 3, 6, 7, 8, 9, 10, 17, etc.). Numerous devices have also been developed to test or assess the development of differentiation in young children. It has frequently been observed that in psychometric tests—e.g., rating scales—the direction and location of responses are partly a function of the scales used, objects rated, as well as the checking style of the subject (e.g., 11). The intent of this paper is to study the relationship between the age and sex of the respondent and checking style on a rating scale—i.e., the semantic differential—in two independent samples of societies which differ greatly in language and/or social systems from the modern western society.

The semantic differential can be used with young children and has been shown to be a reliable instrument (e.g., 4, 5, 12). Osgood, Suci, and Tannenbaum (13) have demonstrated that scale checking styles on semantic differential scales are influenced by a number of variables: e.g., intelligence, age, anxiety, mental disorder of the respondents. It is also a well known fact that a younger child is less capable of differentiating between complex stimuli than an older child or adult. On the basis of Werner's principle of differentiation, it may be stated that a younger child shows relatively less cognitive differen-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on May 20, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Data for this study were collected in connection with two larger studies which were supported by grants from Canada Council.

tiation than an older child. Among several indices that may be used to study greater or lesser cognitive differentiation, scale checking style may be considered to be one of the indices. Therefore, it may be hypothesized that younger children will tend to respond more often on the extreme ends of a rating scale (semantic differential) than the older children. The older children, on the other hand, will tend to respond more often on the middle position or neutral position of the scales than will the younger children. It may also be assumed that early social orientation may influence differently the cognitive styles of boys and girls. It is assumed, further, that the relations observed in non-American societies between age and sex of the respondents and their scale checking styles would be similar to that already observed in the U.S. samples.

B. METHOD

The data reported here were obtained as part of large cross-cultural investigations on the social role perceptions and differentiations among children (see 14, 15).

The semantic differential scales (form II) were used, as described by Osgood *et al.* (13). In order to minimize error variance, role concepts, scales, and method of administration were kept constant in two independent samples. One of the samples consisted of *S*s at two discrete age levels: i.e., eight- and 12-year-old children (Indian sample); the second sample consisted of *S*s on a continuous age range of eight to 13-year-old children (Hutterite sample). The five-point scales were used by all the *S*s.

1. *Subjects*

Two independent samples of children were selected from North-Central India and a communal society in the Canadian Prairies. The sample from the northern provinces of India consisted of 430 Hindi-speaking children who were eight and 12 years old at the time of testing. Because of the fact that intelligence tests in Indian schools are not used widely, children's *IQ* estimates were based on teachers' ratings and *S*s' academic records. The *S*s with extremely low or high *IQ*s were eliminated so that such extreme differences in *IQ*s would not unnecessarily bias children's responses on the semantic scales. Information regarding the father's occupation and income provided the basis for estimates of socioeconomic status in the Indian sample. The majority of children belonged to the Indian middle class.

The Canadian Prairie sample came from 20 Hutterite colonies in rural Manitoba. The *S*s were eight to 13-year-old boys and girls currently enrolled in the public schools in Hutterite communal colonies. Three hundred and fifty-nine *S*s were selected on the basis of their school achievement records and teachers' ratings. The children ranged in their abilities from average to somewhat above-average. In order to control somewhat for the effects of extreme differences in intelligence on scale checking style, very dull and outstanding students were eliminated from the sample. Socioeconomically, the Hutterite colonies form a homogeneous social group located in rural areas.

2. *Semantic Differential Scales*

A set of 15 bipolar scales were selected which tended to have high factor loadings on at least one of the evaluative (E), potency (P), and activity (A) dimensions, as reported by Osgood *et al.* (13). Further, selection of scales was based on relevance, simplicity, representational character, and empirical variance. The scales were subjected to a series of successive empirical tests before the final selection was made. Care was taken to alternate the polarity direction of scales so as to prevent formation of positive preferences. Hindi equivalents of the English words in the semantic scales were obtained from a number of independent bilingual judges in India. The judges were asked first to provide the English equivalent Hindi qualifiers and, next, to give their opinion on the Hindi qualifiers provided by others. The Hindi qualifiers which received the highest degree of consensus as being the best equivalents to the corresponding English words were finally selected.

The scales so selected were as follows: weak-strong, important-common, slow-fast, interesting-boring, afraid-brave, busy-lazy, dull-bright, useful-useless, sad-pleasant, soft-hard, unfriendly-friendly, gentle-rough, dirty-clean, active-quiet, small-large.

3. *Concepts*

The social role concepts were as follows: (*a*) Family roles (males): grandfathers, fathers, brothers; (*b*) Family roles (females): grandmothers, mothers, sisters; (*c*) Occupational roles (professional): doctors, teachers, presidents; (*d*) Occupational roles (services): janitors, waitresses, cowboys; (*e*) Self-Concept: me; (*f*) Friends: boy-friends, girl-friends.

The bilingual judges also offered the concept role equivalents in Indian culture. In the Hutterite communal sample, the following adaptations were made for the occupational role concepts: (*a*) Occupational roles (males): farm-foremen, pigmen, doctors; (*b*) Occupational roles (females); kitchen-women, garden-women, kindergarten teachers.

TABLE 1

INDIAN CHILDREN'S RESPONSE CHECKING STYLE: MEANS AND STANDARD-DEVIATIONS OF EXTREME—1 AND 5—MIDDLE—2 AND 4—AND NEUTRAL—3—SCALE POSITIONS

Subgroups	Age	*N*	Extreme 1 & 5	$\bar{X}$	*SD*	Middle 2 & 4	$\bar{X}$	*SD*	Neutral 3	$\bar{X}$	*SD*
YG I	8	99	17,608	177.9	53.7	3,372	34.1	49.7	1,086	10.9	22.5
YB II	8	94	15,849	168.6	55.8	4,715	50.1	49.5	604	6.4	8.7
OG III	12	108	15,432	142.9	52.4	8,247	76.4	50.3	401	3.7	10.2
OB IV	12	129	15,767	122.2	58.1	12,440	96.4	57.2	242	1.9	11.6

TABLE 2

HUTTERITE CHILDREN'S RESPONSE CHECKING STYLE: MEANS AND STANDARD-DEVIATIONS OF EXTREME—1 AND 5—MIDDLE—2 AND 4—AND NEUTRAL—3—SCALE POSITIONS

Subgroups	Age	*N*	Extreme 1 & 5	$\bar{X}$	*SD*	Middle 2 & 4	$\bar{X}$	*SD*	Neutral 3	$\bar{X}$	*SD*
YG I	8	28	5061	180.75	60.6	1349	48.77	44.0	308	11.0	20.6
	9	34	6078	178.76	59.2	1755	51.61	44.7	327	9.6	19.4
	10	39	6586	168.80	59.5	2310	59.20	47.7	466	11.9	18.5
Total		101									
YB II	8	19	3508	184.63	58.3	925	48.68	40.2	126	6.6	17.0
	9	32	5696	177.93	61.3	1718	53.68	51.3	272	8.5	15.0
	10	27	3959	146.61	62.2	2059	76.26	49.0	362	13.1	23.2
Total		78									
OG III	11	29	4308	148.50	77.1	2392	82.41	60.8	314	10.8	18.8
	12	27	4133	153.07	65.1	1873	69.37	59.5	477	17.6	21.4
	13	36	4989	138.58	63.0	3101	86.14	59.5	565	15.7	21.7
Total		92									
OB IV	11	38	5562	146.37	66.7	3292	86.63	60.2	202	8.2	14.2
	12	27	3386	125.41	69.2	2767	102.48	64.3	287	7.6	10.0
	13	23	3127	135.96	71.9	2281	99.17	61.4	328	12.1	15.2
Total		88									

C. RESULTS

Responses were scored for extreme—1 and 5—middle—2 and 4—and neutral—3—scale positions. These scores were summed for all the adjectival pairs over all concepts to provide "style" scores of an individual respondent. Summing overall subjects' three separate style scores provided the group scores of extreme (E), middle (M), and neutral (N) scale positions. Tables 1 and 2 indicate the checking style scores for the Indian and Hutterite samples respectively.

TABLE 3
ANALYSIS OF VARIANCE: INDIAN CHILDREN'S RESPONSE CHECKING STYLE
(Ages eight and 12 years)

Source	*df*	*SS*	*MS*	*F*	*p*
Between	429	3,065.81			
Sex	1	57.84	57.84	8.61	.05
Age	1	145.85	145.85	21.74	.01
Sex × Age	1	4.04	4.04	.60	NS
Error (bet.)	426	2,858.08	6.71		
Within	860	131,566.50			
SCORE	2	88,277.88	44,138.94	1,051.18	.001
SCO × Sex	2	1,058.49	529.25	12.60	.01
SCO × Age	2	6,292.09	3,146.04	74.92	.01
SCO × Sex × Age	2	161.27	80.63	1.92	NS
Error (within)	852	35,776.77	41.99		

The trend is clear in both the tables, indicating that younger-age children check the extreme scale positions relatively more often than they check the middle or neutral positions. This tendency seems to change steadily (Table 2) with age. As children grow older, they tend to check increasingly the middle positions. For the two age groups of eight- and 12-year-old boys and girls, there is an almost exact replication of response checking scores in the two different and independent samples. This clear trend was tested for significance by a complex analysis of variance design. The raw scores were transformed by the square root formula $\sqrt{X +}\ \sqrt{X + 1}$. The variables in the analysis of variance consisted of three style scores × two sexes × two ages (or six ages for the Hutterite sample). The summary tables employing this complex analysis are presented in Tables 3 and 4. The differences between three style scores are found to be highly significant at $p < .001$ level. There is also a significant score × age interaction ($p < .01$) in both the samples. Thus, so far, influence of age on response checking style is found to be significant in two widely different social-cultural groups.

In the case of the Indian sample, there is also a significant interaction effect

($p < .01$) between score and sex. Such an effect is missing in the Hutterite sample. Moreover, there are significant age difference effects ($p < .01$) between the two discrete age groups of eight and 12 years in the Indian sample. There is also a somewhat significant sex effect ($p < .05$) among Indian respondents.

TABLE 4
ANALYSIS OF VARIANCE: HUTTERITE CHILDREN'S RESPONSE CHECKING STYLE
(Ages eight years to 13 years)

Source	*df*	*SS*	*MS*	*F*	*p*
Between	358	6,309.34			
Sex	1	7.18	7.18	.41	NS
Age	5	243.55	48.71	2.79	NS
Sex × Age	5	6.69	1.34	.007	NS
Error (bet.)	347	6,051.92	17.44		
Within	718	106,477.69			
SCORE	2	67,973.81	33,986.91	665.36	.001
SCO × Sex	2	257.17	128.58	2.52	NS
SCO × Age	10	2,518.58	251.86	4.93	.01
SCO × Sex × Age	10	279.66	27.97	.55	NS
Error (within)	694	35,448.47	51.08		

D. DISCUSSION

The results support the stated hypothesis that older children make increasingly greater use of middle scale positions, while the younger ones check the extreme positions more often than do the older ones. Thus, there is evidence of increased cognitive differentiation with increase in developmental age. It is clearly understood here that age *per se* is only a convenient yardstick to mark the developmental changes in cognitive differentiation. To this extent, Werner's (16) hypothesis is well substantiated that there is a progression from lesser differentiation to increased differentiation in cognitive growth. It is all the more remarkable that these developmental cognitive differentiations could be studied in widely different culture groups by means of the same psychometric instrument (semantic differential).

Light and Zax (11) reported highly similar changes in response checking style with changes in developmental ages of children. Osgood *et al.* (13) have offered further evidences of the same sort. To this extent, the third hypothesis is also supported that changes in response checking style with age are similar in Indian, Hutterite, and American subjects studied. The second hypothesis pertained to the influence of sex differences on response checking styles. The hypothesis is partly supported insofar as Indian respondents show a significant difference in this regard, whereas the Hutterite respondents do not show such significant differences. Here, it is interesting to note that in a number of studies dealing with American subjects, the sex differences were found to be significant (e.g., 1, 13), and in some others they were not significant (e.g., 11). It may be mentioned here that Osgood *et al.* (13) interpreted the data as evidence for three difficulty levels: i.e., all-or-nothing, "black-white" decision being the easiest, intermediate level of difficulty is the "neither" (neutral) judgment, and the most difficult level is presented by the modified scale positions—e.g., 2 and 4 in five-point scale—which requires respondents to make finely graded judgments. The data presented in this study also seem to be consistent with the Osgood *et al.* interpretation.

E. Summary

A sample of 430 children from India and another sample of 359 Hutterite children from Manitoba were studied to determine the possibility of a relationship between age and sex of the children and the tendency to use more frequently the extreme—1 and 5—or middle—2 and 4—or neutral—3—scale positions of the five-point semantic scales. Frequency counts were made for the number of times a subject made an E, M, or N rating. The group means and *SD* were obtained for E, M, and N ratings and were tested for significant differences by a complex analysis of variance design. The age effects in relation to E, M, or N ratings were found significant in both samples. Significant sex difference effects were obtained in the Indian sample but not in the Hutterite sample. The results are discussed in the context of Werner's hypothesis of cognitive differentiation, as well as the Osgood *et al.* interpretation of three difficulty levels involved in the judgmental process.

References

1. Berg, I. A., & Collier, J. S. Personality and group differences in extreme response sets. *Educ. & Psychol. Meas.*, 1953, **13**, 164-169.
2. Bieri, J. Parental identification, acceptance of authority and within-sex differences in cognitive behavior. *J. Abn. & Soc. Psychol.*, 1960, **60**, 76-79.
3. Bing, E. Effect of childrearing practices on development of differential cognitive abilities. *Child Devel.*, 1963, **34**, 631-648.
4. DiVesta, F. J., & Dick, W. The test-retest reliability of children's ratings on the semantic differential. *Educ. & Psychol. Meas.*, 1966, **26**, 605-617.
5. Downing, R. W., Moed, G., & Wright, B. W. Studies of disability: a technique for the measurement of psychological effects. *Child Devel.*, 1961, **32**, 562-575.
6. Dyk, R. B., & Witkin, H. A. Family experiences related to the development of differentiation in children. *Child Devel.*, 1965, **36**, 21-54.
7. Goodenough, D. R., & Eagle, C. J. A modification of the embedded-figures tests for use with young children. *J. Genet. Psychol.*, 1963, **103**, 67-74.
8. Hess, R. D., & Shipman, V. C. Early experience and the socialization of cognitive modes in children. *Child Devel.*, 1965, **36**, 869-886.

9. KAGAN, J., MOSS, H. A., & SIEGEL, I. E. Psychological significance of styles of conceptualization. *Monog. Soc. Res. Child Devel.*, 1963, **28** (No. 2, Serial No. 86), 73-112.
10. KARP, S. A. Field dependence and overcoming embeddedness. *J. Consult. Psychol.*, 1963, **27**, 294-302.
11. LIGHT, C. S., ZAX, M., & GARDINER, D. H. Relationship of age, sex, and intelligence to extreme response style. *J. Personal & Soc. Psychol.*, 1965, **2**, 907-909.
12. MALTZ, H. E. Ontogenetic changes in the meaning of concepts as measured by the semantic differential. *Child Devel.*, 1963, **34**, 667-674.
13. OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. The Measurement of Meaning. Urbana, Ill.: Univ. Illinois Press, 1957.
14. SCHLUDERMANN, S., & SCHLUDERMANN, E. A comparative factorial analysis of semantic structures in monolingual Indian and American school children. *Can. J. Psychol.*, 1969, **23**, 79-93.
15. ———. Social role perceptions of children in Hutterite communal society. *J. of Psychol.*, 1969, **72**, 183-188.
16. WERNER, H. Comparative Psychology of Mental Development. New York: Follet, 1948.
17. WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., & KARP, S. A. Psychological Differentiation: Studies of Development. New York: Wiley, 1962.

Department of Psychology
The University of Manitoba
Winnipeg, Canada

Published as a separate and in *The Journal of Psychology,* 1969, **72**, 263-267.

A COMPARISON OF VOLUNTEERS AND NONVOLUNTEERS FOR ASTRONAUT TRAINING*[1]

Department of Psychology, Ohio University and Naval Aerospace Medical Institute

L. K. WATERS AND ROSALIE K. AMBLER

A. INTRODUCTION

Many projects involving performance in hostile environments utilize volunteer personnel. Studies by Ambler, Berkshire, and O'Connor (2), Bair and Gallagher (3), and Gunderson (5) suggest volunteers for such projects are superior to nonvolunteers on several measures of ability, performance, and motivation. However, for continuing programs it is necessary to maintain a rate of volunteering and volunteer quality over an extended period of time to provide an adequate input of personnel to the program.

Since 1959 all naval aviation students successfully completing the preflight course have received a form offering them the opportunity to volunteer for future duties involving space flights. They are instructed to return the form within one week if they wish to volunteer. No guarantee is given that they will be selected for such duties, and they must confirm their volunteering at further specified points in training if they wish to continue to be considered as a volunteer. Early in the project, Ambler, Berkshire, and O'Connor (2) compared 335 volunteers and 914 nonvolunteers (pooled across all types of flight students) on scores on aptitude tests taken for entry into the program, preflight course grades, peer ratings of officer potential, and rate of attrition from flight training. The volunteer group was superior on all measures.

The present study was a follow-up investigation to determine (*a*) if differences between volunteers and nonvolunteers continued to exist after the project had been in operation for approximately four years, and (*b*) if any differences found were consistent across the various types of students undergoing training.

B. METHOD

The data for this study were based on the records of 2252 students (990 officer and 1262 nonofficer) who completed preflight training during a sin-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on May 26, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Opinions or conclusions contained in the report are those of the authors. They are not to be construed as necessarily reflecting the views or the endorsement of the Navy Department.

gle year. The data were divided into six groups as follows: *Nonofficer:* 1) Naval and Marine Aviation Cadets (commissioned at end of flight training), 2) Aviation Officer Candidates (AOC's are college graduates who have not participated in NROTC; they are commissioned at the end of the 16 week preflight training course); *Officers under Instruction (OI's):* 3) Naval Academy Graduates, 4) NROTC Graduates, 5) Marine Officers, 6) "Other" Officers (procured from other sources—e.g., Newport Officer Candidate School, Coast Guard, Merchant Marine Academy).

For each group, volunteering information, screening test scores, academic grades, and data on attrition from flight training were obtained. In addition, peer ratings on military leadership potential were obtained for the nonofficer groups.

C. Results and Discussion

Table 1 presents the percentages of volunteers in the various groups. There was a marked difference between the officer and nonofficer groups in volunteering rate. Officers volunteered significantly ($p < .01$) more than nonofficers; Naval Academy graduates showed the highest rate of volunteering, while Aviation Officer Candidates (AOC's) had the lowest rate. These two groups were also high and low in the Ambler, Berkshire, and O'Connor (2) study. The greatest change in volunteering rate compared to the earlier study was the increased rate for NROTC graduates (24 percent to 33 percent) and Marine officers (25 percent to 39 percent). The overall volunteering rate was almost identical to that found earlier (27 percent *vs.* 26 percent).

Table 2 gives the means and standard deviations of scores made by as-

TABLE 1
PERCENTAGE OF NAVAL FLIGHT STUDENTS VOLUNTEERING FOR POTENTIAL ASTRONAUT DUTY

Student type	Number	Volunteers	% Volunteers
Nonofficer			
Cadet	723	163	22.6
Aviation Officer Candidate	539	92	17.1
Total Nonofficer	1262	255	20.2
Officer			
Naval Academy Graduate	203	82	40.4
NROTC Graduate	249	83	33.3
Marine Officer	366	141	38.5
Other	172	32	18.6
Total Officer	990	338	34.1
Total combined	2252	593	26.3

TABLE 2

MEAN COMPARISONS OF VOLUNTEERS AND NONVOLUNTEERS ON SCREENING TEST SCORES

Student type	Variable	Volunteers		Nonvolunteers	
		$\overline{X}$	*SD*	$\overline{X}$	*SD*
Cadets	AQT	5.89	1.49	5.76	1.53
	FAR	6.58	1.30	6.37*	1.24
AOC	AQT	6.47	1.59	5.89**	1.51
	FAR	6.67	1.31	6.18**	1.20
OI Academy	AQT	7.61	1.27	7.13*	1.28
	FAR	6.15	1.49	5.50**	1.56
OI NROTC	AQT	7.66	1.25	7.28*	1.38
	FAR	6.84	1.47	6.40*	1.69
OI Marine	AQT	6.78	1.52	6.01**	1.47
	FAR	6.37	1.52	6.09	1.34
OI Other	AQT	6.47	1.45	6.49	1.48
	FAR	5.80	1.47	6.14	1.46

* Difference between volunteers and nonvolunteers significant at $p < .05$.

** Difference between volunteers and nonvolunteers significant at $p < .01$.

tronaut volunteers and nonvolunteers on the Aviation Qualification Test (AQT) and the Flight Aptitude Rating (FAR). These are tests taken at the time of application for naval air training. The AQT is a test of general ability or intelligence, and the FAR is a composite score derived from tests of mechanical ability, spatial ability, and biographical data. The AQT has been shown to be related to ground school grades, while the FAR has shown predictive validity for flight training performance (1). Volunteers for astronaut duty scored significantly ($p < .05$) higher than nonvolunteers in eight of 12 comparisons.

The first row for each student type in Table 3 presents the data for the preflight final grade. Volunteer means were significantly ($p < .05$) higher than nonvolunteer means for all student types. Peer rating means and standard deviations are given in the second row of the Cadet and AOC groups. This variable is a score based on student nominations of their classmates for leadership ability and has demonstrated validity for predicting future pilot performance (4). The mean peer rating was significantly ($p < .05$) higher for volunteers in both the Cadet and AOC groups.

Table 4 shows the attrition from flight training for the volunteer and nonvolunteer groups. Preflight attrition cases were not included, since these students did not have an opportunity to volunteer. In all except the Cadet group, volunteers had smaller attrition rates than nonvolunteers. No reason

TABLE 3
MEAN COMPARISONS OF VOLUNTEERS AND NONVOLUNTEERS ON PREFLIGHT PERFORMANCE

Student type	Variable	Volunteers $\overline{X}$	Volunteers *SD*	Nonvolunteers $\overline{X}$	Nonvolunteers *SD*
Cadets	Preflight final	51.20	4.72	50.28*	4.67
	Peer rating	51.90	10.53	50.18*	9.72
AOC	Preflight final	53.61	4.84	50.90**	4.73
	Peer rating	55.37	9.93	51.09**	9.56
OI Academy	Preflight final	57.18	4.97	54.62**	4.96
OI NROTC	Preflight final	58.60	4.18	55.92**	5.09
OI Marine	Preflight final	55.45	5.52	50.31**	5.65
OI Other	Preflight final	54.16	4.85	47.50**	8.32

* Difference between volunteers and nonvolunteers significant at $p < .05$.
** Difference between volunteers and nonvolunteers significant at $p < .01$.

for the marked difference between cadets and the other groups in the attrition rates of volunteers and nonvolunteers is available.

In general, volunteers for astronaut duty were superior to nonvolunteers on measures of ability, early performance in training, and completion rate. These data support the results of Ambler, Berkshire, and O'Connor (2) obtained during the early phases of the volunteering project and indicate that volunteering rate and volunteer quality are maintained over extended time periods.

D. SUMMARY

After a program of offering flight students the opportunity to volunteer for astronaut training had been in effect several years, volunteers and non-

TABLE 4
ATTRITIONS FROM FLIGHT TRAINING FOR VOLUNTEERS AND NONVOLUNTEERS

Student type	Volunteers *N*	Volunteers %	Nonvolunteers *N*	Nonvolunteers %
Cadet	52	31.9	140	25.0**
Aviation Officer Candidate	23	25.0	153	34.2**
Naval Academy Graduate	13	15.8	20	16.5
NROTC Graduate	16	19.2	36	21.4
Marine Officer	24	17.0	52	23.1**
Other Officer	9	28.1	46	32.9
Total	117	23.1	47	27.0

** Difference between percentage attritions for volunteers and nonvolunteers significant at $p < .01$.

volunteers were compared on selection test scores, preflight performance, and attrition from flight training to determine if volunteer quality was maintained. In general, volunteers were higher on ability and preflight performance measures than nonvolunteers. Attrition data were inconsistent. Volunteering rate was the same as that found during the initial stages of the program.

REFERENCES

1. AMBLER, R. K. Characteristics of the revised naval aviation selection tests administered under experimental conditions. Report No. NSAM-234, Naval School of Aviation Medicine, Pensacola, Florida, 1955.
2. AMBLER, R. K., BERKSHIRE, J. R., & O'CONNOR, W. F. The identification of potential astronauts. Report No. NSAM-785, Naval School of Aviation Medicine, Pensacola, Florida, 1961.
3. BAIR, J. T., & GALLAGHER, T. J. Volunteering for extra-hazardous duty. *J. Appl. Psychol.,* 1960, **44**, 329-331.
4. BERKSHIRE, J. R., & NELSON, P. D. Leadership peer ratings related to subsequent proficiency in training and in the fleet. Special Report 58-20, Naval School of Aviation Medicine, Pensacola, Florida, 1958.
5. GUNDERSON, E. Performance evaluations of Antarctic volunteers. Report No. NMRU 64-19, Medical Neuropsychiatric Research Unit, San Diego, California, 1964.

Dept. of Psychology
Ohio University
Athens, Ohio 45701

Published as a separate and in *The Journal of Psychology*, 1969, **72**, 269-292.

QUANTITATIVE BEHAVIOR OF WARRING NATIONS*[1]

Department of Psychology, Stanford University

JOHN VOEVODSKY[2]

A. INTRODUCTION

Warfare is demonstrated to be an orderly and repetitive behavioral event. During the last 100 years, the United States has fought four major wars to conclusion: The Civil War, World War I, World War II, and the Korean War. For the past eight years, we have been engaged in a fifth and not yet concluded war, the war in Vietnam, which has occupied our minds, our troops, and our resources longer than any other war in our history.

And yet, despite its record duration, the Vietnam War appears to be progressing, as of this writing, in the same orderly manner as our four previous wars. An inquiry into the repetitive behavioral patterns of nations at war, particularly of the United States, our allies, and our enemies during the last 100 years, reveals that we and the enemy are acting today in the same way as we have acted in the past. Because of this orderliness of our behavior, it now appears that we may be approaching a crisis point in the Vietnam War where either a settlement is possible or another major escalation is indicated.

This orderly manner in which the Vietnam War and the previous wars have progressed is based on an examination of the phenomena—the vital statistics—of modern armed conflict by means of cybernetic analysis. Thus, it is an identification and demonstration of the orderly progress of warfare that is the primary concern of this paper, together with the derivation of equations that express mathematically the relationships among the vital statistics of warring nations.

---

* Recommended by Ernest R. Hilgard of the Editorial Board, and received in the Editorial Office, Provincetown, Massachusetts, on May 26, 1969, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The author expresses appreciation for ideas and encouragement from Mr. Sidney Burr Brinckerhoff, Fellow, Company of Military Historians; and Professors Ronald Grant (Physiology), Ernest R. Hilgard (Psychology), Krishnamurty Karamcheti (Aeronautics and Astronautics), James McGregor (Mathematics), Robert C. North (Political Sciences, Studies of International Conflict and Integration) of Stanford University; and Dr. Samuel C. McIntosh, Research Associate, Aeronautics and Astronautics, Stanford University; and Dr. S. L. Waleszczak, Chief Scientist and RADM E. R. Zumwalt, Jr., Director, Systems Analysis Division, Office of the Chief of Naval Operations. For all interpretations of the data, however, I am solely responsible.

[2] The author is a behavioral cyberneticist with the Cybernetica Research Corporation, P.O. Box 7862, Stanford, California, 94305. This article is based on research performed while a Research Associate at Stanford University and while a consultant to the Office of the Chief of Naval Operations.

The present investigation is purely of an exploratory nature, it should be emphasized; and no pretense is made that it is possible at present to explain fully and in detail all the factors which result in the orderly progress of warfare.

All such endeavors, though, however bravely undertaken and whatever their results may be, have a theoretical origin. This research is no exception. In fact, it has two origins: one in the quantitative behavioral concepts of the psychologist S. S. Stevens (9) and the other in the cybernetic concepts of the mathematician Norbert Wiener (12). In essence, Stevens shows that quantification of behavior is readily achieved by the assignment of numerals to distinguishable behavioral events in the career of living systems. These behavioral events become the system variables, which are time variant. The system variables, in addition, appear to have the characteristics inherent in linear second-order feedback-control systems of the type first conceived by Wiener. Furthermore, the work on the quantitative nature of warfare performed by L. F. Richardson (8) and F. W. Lanchester (5) supports these general observations.

B. The Vital Statistics of War

The quantitative behavioral data which best describe modern warfare fall into three categories: battle strength, battle casualties, and battle deaths. By coincidence, these are the vital statistics of war, because nations habitually count their troops, assess their attrition, and record the results of combat for current and historical uses. From the point of view of the analyst, though, these particular items of data have a value in addition to their ubiquity. They have an advantage over other war statistics, because they appear to be remarkably stable and accurate, since in most cases they have stood the scrutiny of countless researchers over long periods of time.

It should be further emphasized that only official government sources are used for the statistics in this analysis, with two exceptions. French and German battle deaths and casualties during World War I were taken from Winston S. Churchill's compilation of official records (2). Union and Confederate statistics for the Civil War were taken from Thomas L. Livermore's compilation of official records (6).

The basic vital statistics, battle strength, battle casualties, and battle deaths, unless carefully defined, may mean different things to different analysts. In this analysis, the numbers used for all wars represent only Army personnel, officers and enlisted men, reported at the specified instant of time that is under consideration. This applies whether the troops are Union, Confederate, French, German, British, or American. In the case of the Vietnam War, however, numbers of total Department of Defense forces, which includes Army, Air Force, Navy, Marine Corps, and Coast Guard, are used separately and in addition to the numbers used for the Army personnel.

For the United States Army, battle strength is the total number of officers and enlisted men at any instant, *t*, deployed in the battle zone. For the Vietnam War, the battle zone is considered to be South Vietnam, and the number of Army personnel in South Vietnam at any instant is considered to be the Army battle strength. This does not include Army personnel who are off shore or are in any other location, such as Thailand.

Army battle strength for the Korean War is the total number of Army personnel on the Korean Peninsula at any instant (3); for World War II, it is the total number stationed outside the continental limits of the United States at any instant of time (4); for World War I it is the total number in France at any instant (1, p. 15). Similarly, British Army battle strength is the total number of British Army personnel in France at any instant of time during World War I (1, p. 14). Comparative statistics for German Army battle strength during World War I were not available in suitable form.

For the Civil War, the Union Army battle strength is the total number of personnel on the Union rolls at any instant, and the Confederate Army battle strength is the total number of personnel on the Confederate returns at any instant (6, p. 47). Of all the war statistics included in this analysis, the Civil War strength data show the widest fluctuations and are considered the least reliable of all the data.

Army battle casualties for the Vietnam War (15), the Korean War (14), and World War II (13) are the accumulated totals from the beginning of the particular war up to any instant of time, *t*, of all Army personnel killed in action, dead as the result of wounds or injuries received in action, wounded or injured in action, missing in action, captured by the opposing forces, or taken into custody as internees by the authorities of a neutral country. The term "in action" characterizes a casualty status in which an individual suffered death or injury as a direct result of enemy action, either while engaged in a battle, going to it, or returning from it. Excluded from United States Army battle casualties are injuries or deaths that resulted from exposure to the elements or from diseases.

Army battle deaths for the Vietnam War, the Korean War, and World War II are the accumulated totals from the beginning of the wars up to any instant, *t*, of Army personnel who were killed in action, died as a result of injuries or wounds received in action, or declared dead while missing in action.

For the United States Army in World War I, the time variants in the battle casualties are not known. The time variants in battle deaths are known. Total U.S. battle casualties at the end of World War I included dead, wounded, missing, and prisoners (1, pp. 120-122). Similarly, mutually caused British *versus* German (16), and French *versus* German (2, p. 39) battle casualties include dead, wounded, missing, and prisoners (2). Only the U.S. Civil War battle casualties, which included dead and wounded, exclude missing and prisoners because the data were unavailable (6). The interested auditors of this paper should consult the original sources cited for detailed definitions of battle strength, battle casualties, and battle deaths.

The British-German and French-German World War I and U.S. Civil War vital statistics should be considered as approximations. Differences exist in the data for each war, which must be considered by the analyst as measurement error. The size of the error, unfortunately, is not known at this time. Also, when the analyst compares the errors between wars, he should assume that part of the errors are due to differences in accounting procedures. The errors that might occur in the United States data for the Vietnam War, the Korean War, and World War II, however, can be considered the smallest; since statistics for the three wars are compiled by the same government, are close together in time, and major changes in accounting procedures do not appear to have been made. No detailed studies have been made of the exact accounting methods employed by the other governments because of the exploratory nature of this research. The data have been taken more or less on faith from official sources, except where noted. The reader should be assured, however, that when conclusions are drawn, he will be alerted to the possible error in the base statistics so that he may discount the conclusions to the degree he chooses.

C. METHOD OF ANALYSIS

A mere compilation in tabular form of these three items of data, arrayed over time, reveals very little about the progressive nature of the statistics for a particular war, the relationships that exist between opposing forces in a particular war, or the relationships that exist among the statistics of different wars. Yet it is an identification of these significant relationships that is being sought in this analysis through a graphic and mathematical manipulation of the three sets of vital statistics. Thus, here are identified and separated out for examination the relationships between battle casualties and battle deaths, between battle strength and battle casualties, between opposing forces (the effect of their coupled statistics), and between battle strength and time. Other by-products, such as the effectiveness of weapons systems, can also be identified.

In order to accomplish this purpose of identifying significant relationships, the standard scientific method of analysis is followed. The published statistics were gathered and put into the form dictated by the theories of Stevens and Wiener. For example, one important feature of the dictated form is to record accumulated totals over time. Any two sets of these data, thus compiled, define coordinate points. Semilog paper is used to plot the relationship of the statistics to time, and log-log paper is used to plot the relationship of one vital statistic to another. This is done because semilog and log-log paper best illustrate the respective exponential and power relationships there were found to exist. Finally, the data, when compiled in this manner for all five wars, are observed to fit very closely four simple mathematical laws.

The connection of one data point to another on the graph paper, of course, results in two types of lines. The exponential functions produce curves and straight lines. The power functions produce straight lines. In presenting these data graphically a suitable mean line has been drawn through the data points by inspection. Due to the exploratory nature of the research, it is considered unnecessary to employ more rigorous methods of curve fitting.

Thus, it must be said that the basic method, or scheme of analysis, is not materially different from that followed in any other responsible scientific investigation. What is different here are the items chosen for analysis, the uniqueness of their arrangement, and the results of their manipulation.

D. BATTLE DEATHS AND BATTLE CASUALTIES

Identification of the mathematical relationship between battle deaths and battle casualties is the first of the four equations presented here to demonstrate the orderly progress of warfare. The relationship is shown in Figure 1 by a graphic representation of a summary of the United States Army statistics for the Vietnam War, the Korean War, World War II, World War I, and the Civil War (both Union and Confederate sides). In order to avoid a confusion of lines and symbols, however, the British and German statistics for World War I are not shown in Figure 1, even though they portray an identical relationship to the ones actually shown.

This relationship between battle deaths and battle casualties can be described mathematically from an inspection of the plotted curves by a power function, $D = \gamma C^{\delta}$, where D is the battle deaths at any instant, t; C is the battle casualties at the same instant, t; and γ and δ are constants, which as a by-product define the net effectiveness of the enemy's weapons systems over the effectiveness of all countermeasures designed to protect troops. γ and δ define a family of curves, or levels, between the battle deaths and battle casualties for each war. This formulation permits the analyst to view a coalition of nations as one weapon system.

Since the time-variant nature of the United States Army battle casualties for World War I is not known, but the final battle casualties and the time-variation in battle deaths are known, the slope given to the dashed line in

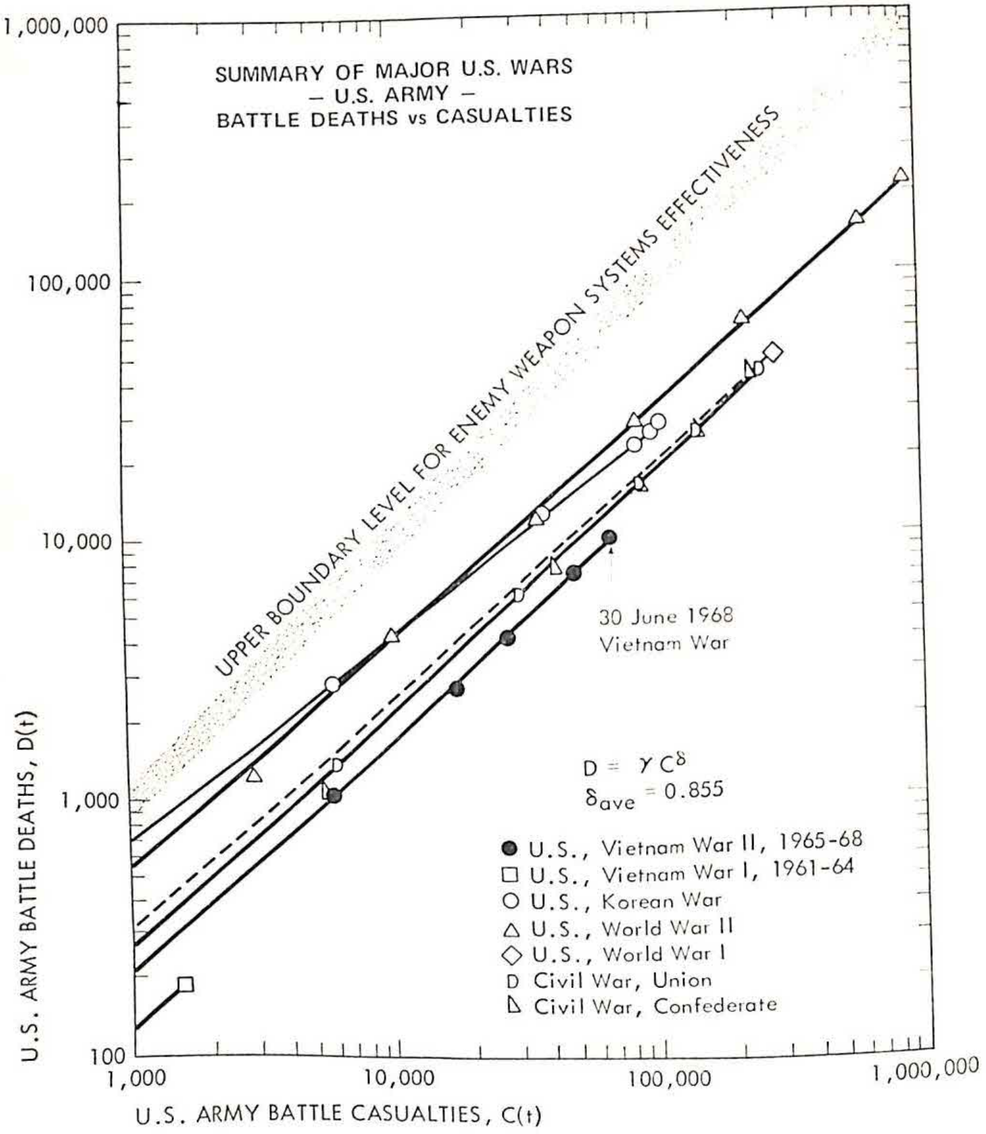


FIGURE 1
SUMMARY OF MAJOR U.S. WARS—U.S. ARMY—BATTLE DEATHS *vs* CASUALTIES

Figure 1 that is drawn from the final casualty point is the average slope, δ_{ave}. This average slope was computed by adding the slopes for the Union and Confederate sides of the Civil War, the United States in World War II, and the United States in the two phases (to be defined later) of the Vietnam War and dividing the sum by five.

Quite automatically, the formulation also establishes the upper boundary of enemy weapons system effectiveness. On the graphs, Figure 1 and Figure 2, this maximum effectiveness boundary represents points where, at a given time, all battle casualties are also battle deaths. Hence, any data point which falls above this boundary represents the physically impossible condition in which more fatalities than casualties occur. The main point here is that the purpose of the weaponry is to destroy the enemy's fighting ability by killing

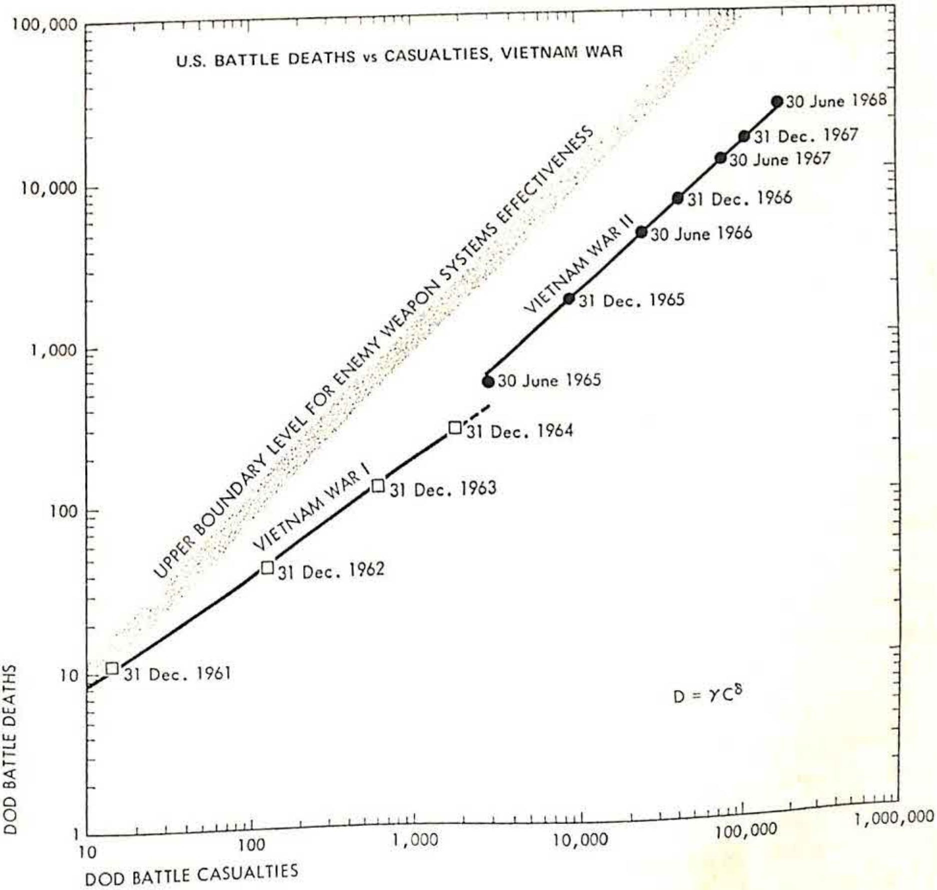


FIGURE 2
U.S. BATTLE DEATHS *vs* CASUALTIES, VIETNAM WAR

his troops. Therefore, the relationship of deaths to casualties is one index of weapons effectiveness. The upper limit of weapons effectiveness, by this definition, is when casualties equal deaths.

In no war has a sudden improvement in enemy weapons system effectiveness been more noticeable than in the Vietnam War, represented separately in Figure 2 in terms of Department of Defense battle deaths to battle casualties. Here, a smoothly progressive and constant relationship is observed to have been maintained from December 31, 1961, through December 31, 1964, which for the sake of convenience is identified as Vietnam War I. In the year 1965, though, a significant change in the relationship occurred, which has held constant through June 30, 1968. This second phase of the war is identified as Vietnam War II.

The enemy weapons system effectiveness level on which Vietnam War I was fought is entirely different from the level on which Vietnam War II is being fought. That these two effectiveness levels are under other major United States war levels is the net result of the effect of the enemy and the countereffect of our modern medical treatment, helicopter evacuation service, and other new protective systems.

These are modern times, though, and this particular enemy theoretically has access to an even more sophisticated weapons system than he is now employing. It thus appears possible that the Vietnam War could suddenly escalate to levels experienced during the Korean War and World War II and become Asian War I merely through the introduction of tanks, air power, missiles, and/or Chinese troops. An escalation in enemy weapons system effectiveness might also occur simultaneously through the employment of Chinese- and Soviet-directed tactics and strategies.

What can happen and what might happen in Vietnam in terms of battle deaths to battle casualties relationships, however, will lie somewhere between the present level and the upper boundary of weapons system effectiveness as shown in Figure 1 and Figure 2. The potential for change in the future, thus defined, lies within a limited range.

E. Battle Strength and Battle Casualties

Identification of the mathematical relationship between battle strength and battle casualties is the second of the four equations presented to demonstrate the orderly progress of warfare. This relationship is shown in Figure 3 by a graphic representation of a summary of United States Army statistics for the Vietnam War, the Korean War, World War II, World War I, and British participation in World War I. Statistics for the Civil War are not included, because the strength figures as a function of time for both sides are known to be of questionable accuracy.

Once again a relationship pattern, one that is common to all wars analyzed, is immediately discernible from an inspection of the plotted curves. This pattern, too, can be described by a power function, $S = \alpha C^{\beta}$, where S is the total Army strength at any instant, t; C is the Army battle casualties at the same instant; α and β are constants, which in combination define the intensity of the fighting in the battle zone. α also represents the theoretical initial battle zone strength at the moment the first battle casualty is sustained. In this sense, α is a battle-preparedness coefficient, and β is the replacement coefficient.

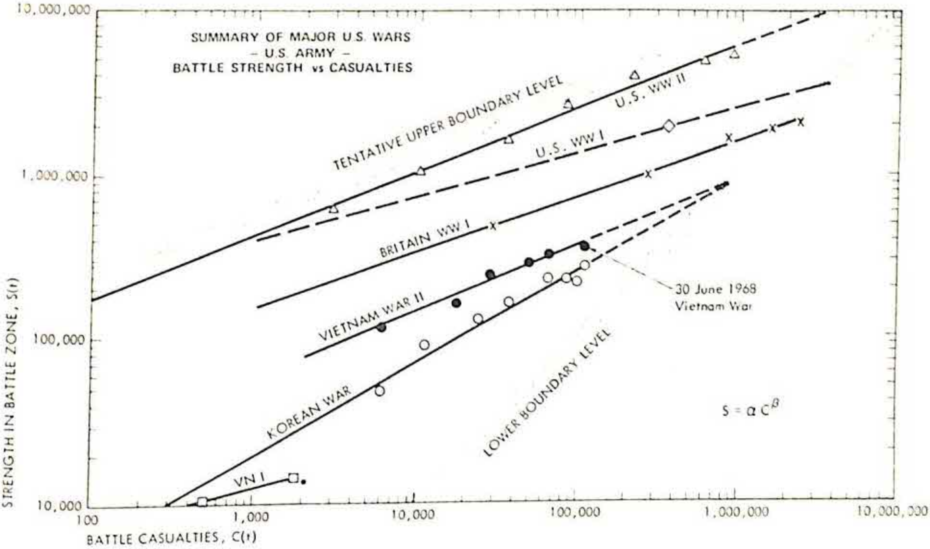


FIGURE 3
SUMMARY OF MAJOR U.S. WARS—U.S. ARMY—BATTLE STRENGTH *vs* CASUALTIES

The equation, however, represents other conditions. It states that the number of Army troops in a battle zone at a given instant is determined by the battle casualties. It also states the converse. In other words, it implies that each new casualty causes the influx of new troops by a predictable amount and, conversely, that new troops produce additional casualties. This is evident, because if a simple mathematical manipulation (the logarithmic derivative) is taken, then $\frac{dS}{S} = \beta \frac{dC}{C}$ results. From Figure 3, β is seen to be less than 1, where β varies very roughly from .57 for the Korean War, .38 for World

War II, .37 for the British experience in World War I, and .36 for Vietnam War I and II. If β were zero, then battle strength would be a constant a, equal to the initial fighting strength. Thus β is, then, the replacement coefficient, where the strength, when doubled in the battle zone, causes the casualties to go up roughly by six times for all wars considered, except for the Korean War, where they went up roughly by four times.

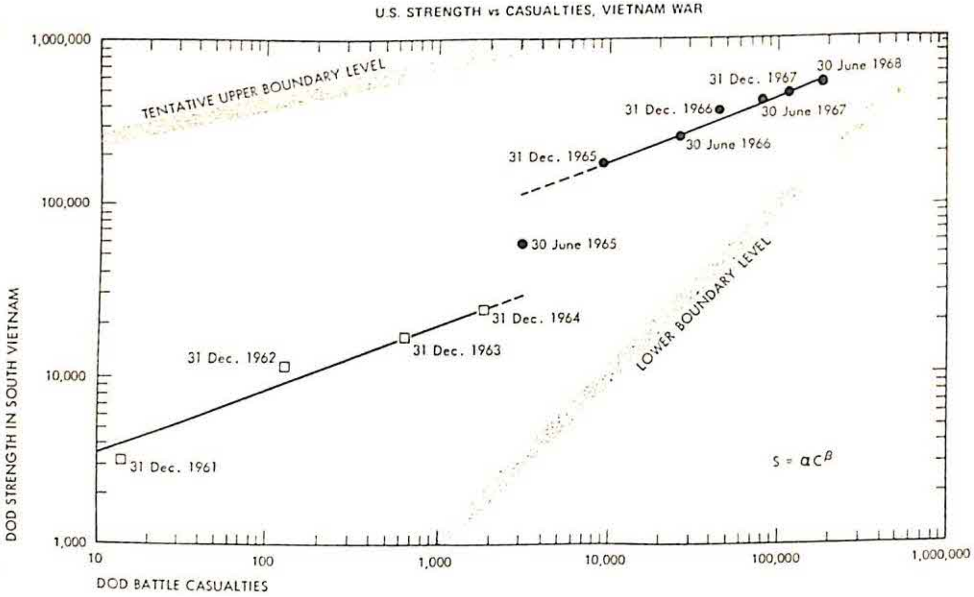


FIGURE 4
U.S. STRENGTH *vs* CASUALTIES, VIETNAM WAR

While it is acknowledged that consideration of five data points, in this case five wars, is not necessarily proof of the existence of a phenomenon, an examination of the relationships between the wars suggests that two important conclusions can be drawn.

The first is that there are limits in strength build-ups and casualties a nation will sustain, beyond which it either accepts defeat, changes its leadership, or acquires new allies. This idea is graphically illustrated in Figure 3 and Figure 4 by a lower boundary level, which defines the points where battle casualties and battle strengths are equal.

This line, in effect, represents a boundary to the right of which a nation cannot seem to pass and still sustain its own war effort. For example, Great Britain at the end of World War I suffered roughly two million battle casualties with a total strength in France at the end of the war of two million soldiers. Since the population of Great Britain at the time was 60 million, this represents 3.33 percent of her 1918 population. It is revealing to note that 3.33 percent of the United States population (110 million during World War I) is 3.67 million combat troops, or casualties, which is the point where the United States experience in World War I intersects the lower boundary when it is extended as shown by the dashed lines in Figure 3.

Furthermore, if the two United States wars in Asia, the Korean War and the Vietnam War, are extended in the same manner, they intersect with each other and with the lower boundary level at the same point, 850,000 casualties and men. Also, the estimated absolute ceiling on the number of men physically fit for active war service for the United States during World War II was between 15 and 16 million (7), which again is the intercept of the extension of the United States World War II with the lower boundary level. This point represents 11.4 percent of the United States population of 131 million in 1940.

The second conclusion that can be drawn from an examination of the relationship of battle strength to battle casualties over five wars is that there appears to be an upper boundary level, or limit on the degree of mobilization that the citizens of a nation will consent to as they prepare for and participate in a major war. This upper level becomes a sloping line when it is represented on the graph, which in this instance is drawn backward from the point where all potential soldiers could become casualties. The angle of the line is drawn according to a slope that is the average for all wars analyzed. This average slope is $\beta_{\text{ave}} = .38$.

If one assumes that, as during World War II, 11.4 percent of the United States population represents the maximum number of soldiers we could prepare for combat, then our present 200 million population would produce a potential reservoir of 22.8 million men for the Vietnam War. The data show, however, that a nation does not mobilize all its potential fighting men at once. It characteristically mobilizes gradually as casualties are incurred. Therefore, this gradual mobilization, mathematically represented by $\beta_{\text{ave}} = .38$, appears to be a behavioral characteristic of warring nations: to maintain a balance between the maintenance of combat-eligible men in their normal productive activities in the society and the complete mobilization and commitment to battle of these same men in an effort to protect the nation's way of life.

Finally, when the strength to casualty relationship is applied to the Vietnam War, as illustrated in Figure 4, it may be clearly seen that a jump in activity occurred in 1965. This was the point at which the United States escalated its strength, expressed mathematically by changing a to a higher level of effort, in response to the change in enemy weapon system effectiveness. For the past two and one-half years the relationship of strength to casualties has not changed materially. Since the data point for June, 1968, is slightly to the left of the mean line, one can conclude that the United States battle strength is being kept down in comparison to the casualties received. This means that slightly more casualties are being inflicted by the enemy than United States strength calls for. Furthermore, since the data point for June, 1968, is so close to the lower boundary level, this second phase of the Vietnam War is shown to be one of the most intensely fought wars the United States has ever experienced.

F. RELATIONSHIP BETWEEN OPPOSING FORCES

Identification of the mathematical relationship between strength to strength, casualties to casualties, and deaths to deaths for opposing armies for a particular war is the third of the four equations presented to demonstrate the orderly progress of warfare. This relationship is shown in Figure 5 for the French *versus* German Armies, and is discernible in Figure 6 for the British *versus* the German Armies during World War I, and Figure 7 for the Union and Confederate Armies during the Civil War. Analyses for the other wars were not made, because data on enemy forces in the Vietnam Wars, and Korean War, and World War II were either not available or were not in suitable form.

Here, too, a discernible relationship can be deduced from the resulting curves, which can be described by a power function, $Y = jX^k$, where X and Y are identical statistics at the same instant of time for opposing forces. The coupling coefficients are j and k. In all cases studied, however, k is found to be almost equal to 1. In fact, k varies between .95 and 1.04. Thus, by setting k equal to 1, the coupling relationship assumes the simpler form of $\frac{dY}{dX} = j$. This equation states that if one side changes its war variables, a similar change occurs on the other side. Lanchester has defined j as the exchange ratio; i.e., the ratio between those combatants lost on one side to those combatants lost on the other. Also, an increase in troops on one side results in an increase in troops on the other. This ratio also is intuitively expected because when two armies come together in combat, firing begins almost simultaneously from both sides. When the combat ends, the firing ends simultaneously on both sides. If a large battle occurs, both sides suffer large losses at the same time.

This is true regardless of whether the battle is classified as a success, a failure, a defeat, a rout, or a victory.

Historically, the Germans have been accused of dividing their casualty statistics in half for political purposes. This may or may not be the case. What is important is that the casualty data for both sides have essentially the same shape when plotted as in Figure 6. The Civil War also shows this expected characteristic of similarity in shape, and even in magnitude for the casualty statistics. Thus the most important point to be realized here is that if the analyst is interested in when a war begins and when it might end, as indicated by a leveling of hostilities (measured by strengths, casualties, and deaths), he can determine this from the casualty data of only one of the two combatants. True, the other side might have a higher or lower strength ratio, or casualty

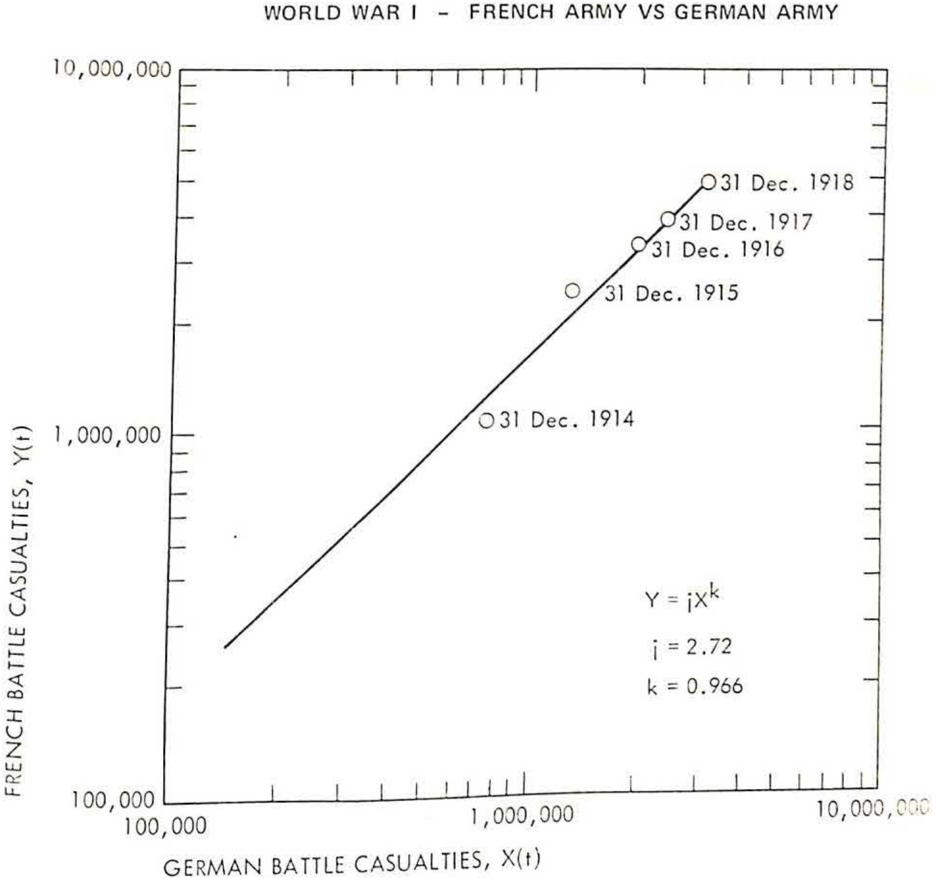


FIGURE 5
World War I—French Army *vs* German Army

ratio, or kill ratio, but the leveling of hostilities occurs simultaneously and in like manner on both sides. The reader should be cautioned, however, that this conclusion is based on observation made of only two wars and within these two wars on only six relationships. They are Union and Confederate strength, Union and Confederate casualties, Union and Confederate deaths, French and German casualties, British and German casualties, and British and German deaths. Thus, on this basis one may conclude that opposing warring nations have similar behavioral patterns over time during a war even though the ratios between statistics may be different.

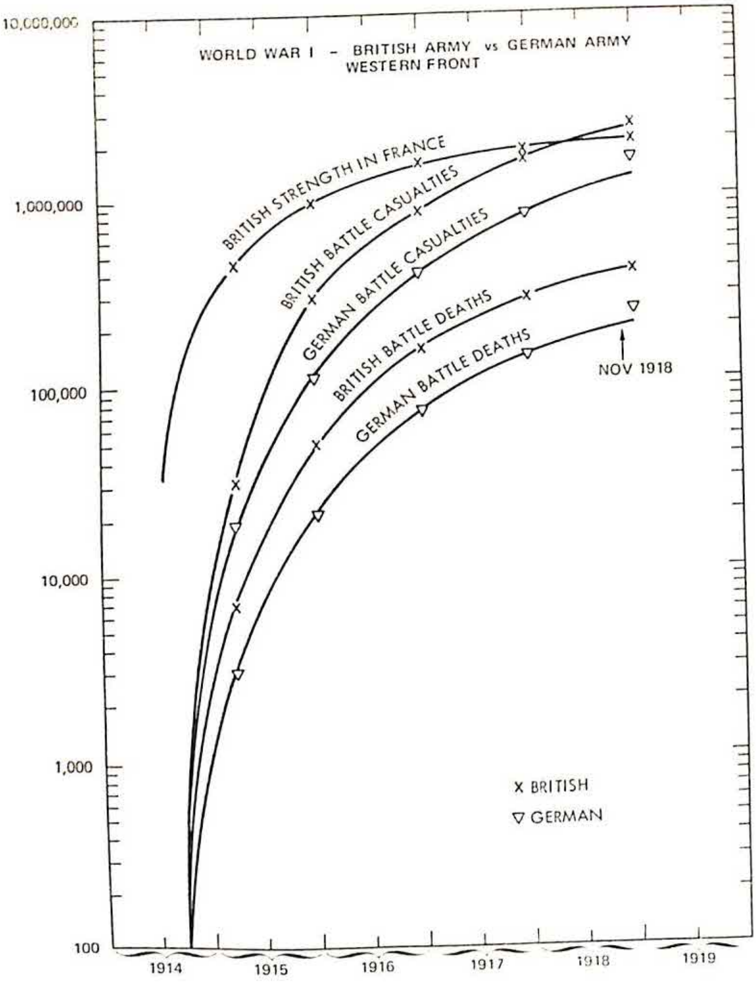


FIGURE 6
World War I—British Army *vs* German Army, Western Front

As this conclusion applies to the Vietnam War, one can say: If the coupling constants do not change, then the leveling of hostilities (as measured by the vital statistics of warfare) is clearly discernible from the United States statistics alone. If, on the other hand, the North Vietnamese escalate the war, the first indication to the United States will be a rise in casualties above projected levels. The coupling coefficients will then have been changed by the enemy and a new and bigger war will have begun.

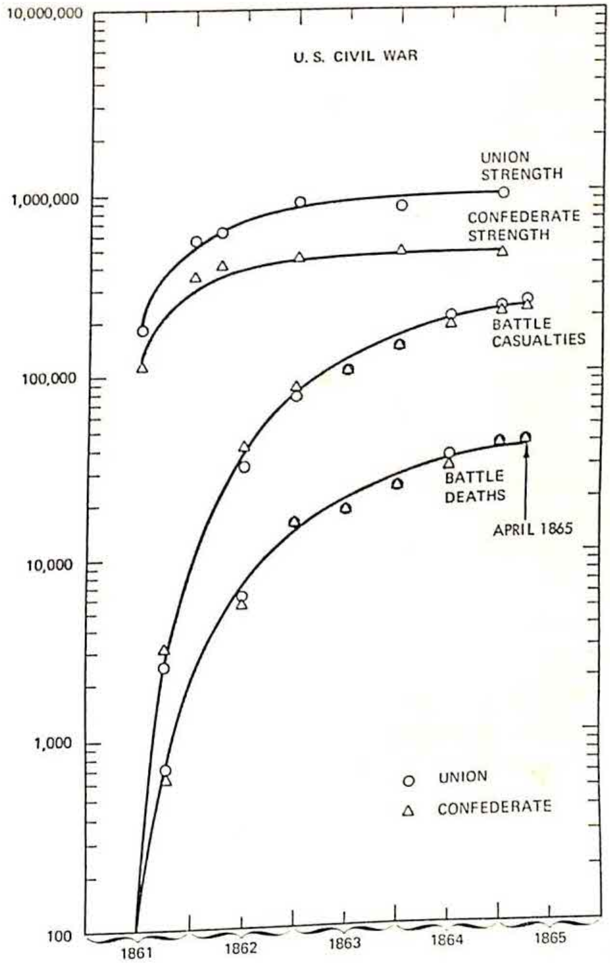


FIGURE 7
U.S. Civil War

G. Battle Strength and Time

Identification of the mathematical relationship between Army battle strength and time is the last of the four equations presented to demonstrate the orderly progress of warfare. Since battle deaths are related to battle casualties and in turn are related to battle strength, the relationship between casualties and between deaths and time can be established by means of algebraic manipulations once the mathematical relationship between battle strength and time is known.

If the behavior of warring nations can be characterized by linear second-order feedback-control systems equations of the type first conceived by Wiener and supported by the work of Richardson, Lanchester, and others (10, 11), then the solutions are exponential in nature.

The general solution for a second-order system is $S = A + Be^{\lambda_1 t} + Ce^{\lambda_2 t}$ where A, B, C, λ_1, and λ_2 are characteristic constants of the system. For the wars studied it is found that the first exponential term quickly disappears, while the strength is less than 10 percent of the ultimate forces. The remaining 90 percent of the ultimate fighting force is characterized by only one time constant. Thus, if the analyst is concerned only with the last 90 percent of a war, first-order differential equations appear to describe accurately the time-variant nature of the vital statistics. Second-order equations appear to be needed only when the beginnings of wars are to be analyzed. Since the Vietnam War is found to be in its terminal stages, the first 10 percent of wars have not been analyzed.

To develop these ideas, one considers first the case of Vietnam War II, shown in Figure 8a and Figure 8b. S denotes, in this case, the total Department of Defense strength in South Vietnam at any instant, t. The variation of battle strength, S, with time, t, is shown in Figure 8a. It is seen that if the conditions which have been acting in a particular manner over the past three years continue in an unchanged manner, the battle strength, S, approaches an upper limit, say S_∞, the ultimate number of forces that are to be engaged. By extrapolation in Figure 8a, one finds that S_∞, to a good approximation, is 600,000 Department of Defense personnel. To obtain a simple analytical representation for the function S, one finds, as shown in Figure 8b, that $\log(S_\infty - S)$ varies linearly with time; i.e., the curve $\log(S_\infty - S)$ *versus* t is a straight line. If $-1/\tau$ denotes the slope of this line and if S_0 denotes S at some initial time, t_0, the equation for the line is

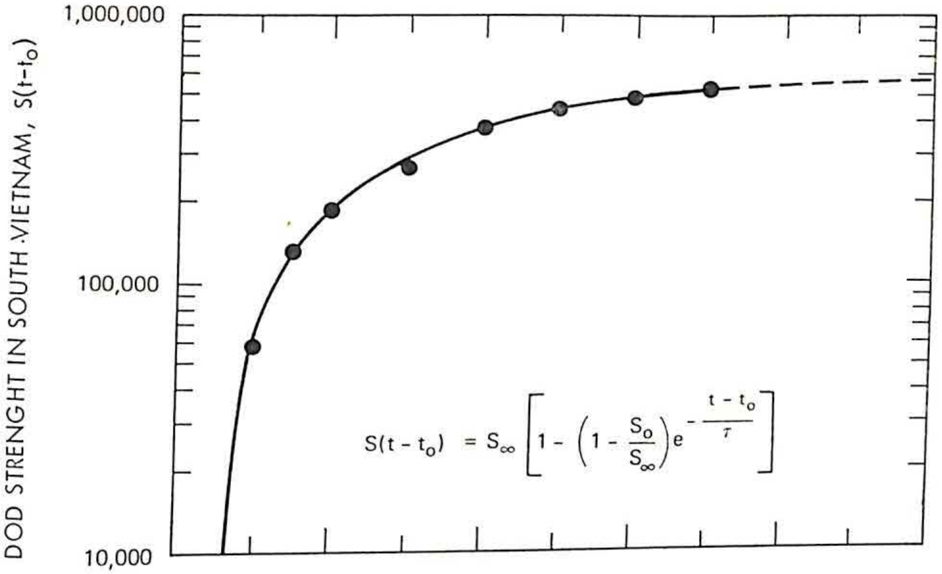
$$\log\left[\frac{S_\infty - S(t - t_0)}{S_\infty - S_0}\right] = -\frac{t - t_0}{\tau}$$

or, equivalently,

$$\frac{S(t - t_0)}{S_\infty} = 1 - \left(1 - \frac{S_0}{S_\infty}\right) e^{-(t - t_0/\tau)}.$$

THE TIME VARIATION IN TOTAL DOD STRENGTH IN SOUTH VIETNAM

a. Strength vs Duration



b. Reciprocal Strength vs Duration

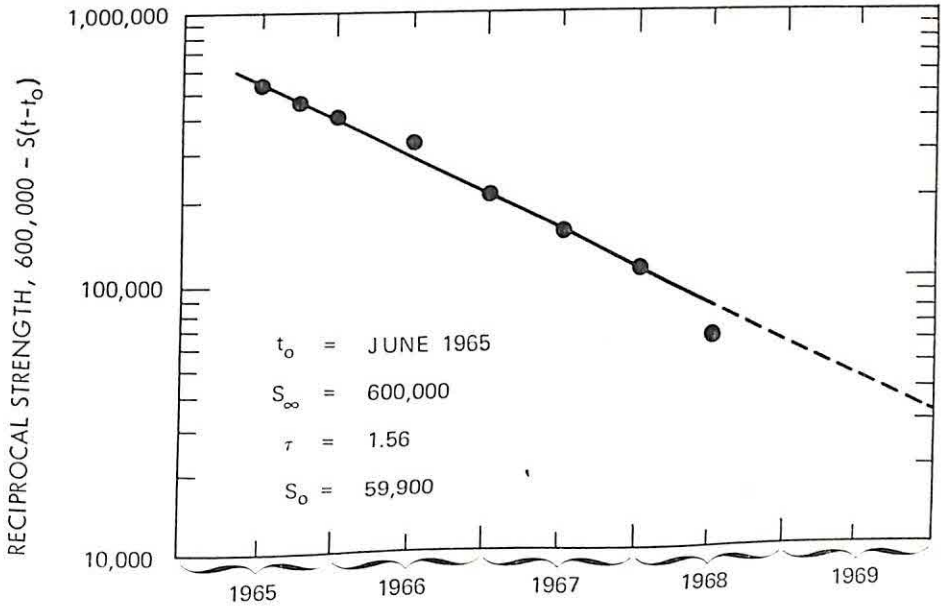


FIGURE 8

The Time Variation in Total Department of Defense Strength in South Vietnam

Generally, for the wars analyzed here the initial strength, S_0, is much less than the final strength, S_∞, at least by a factor of 10. It is, therefore, reasonable to assume that the ratio of initial strength to final strength is much less than one and the above expression may be expressed in the simpler form of

$$\frac{S(t-t_0)}{S_\infty} = 1 - e^{-(t-t_0/\tau)}.$$

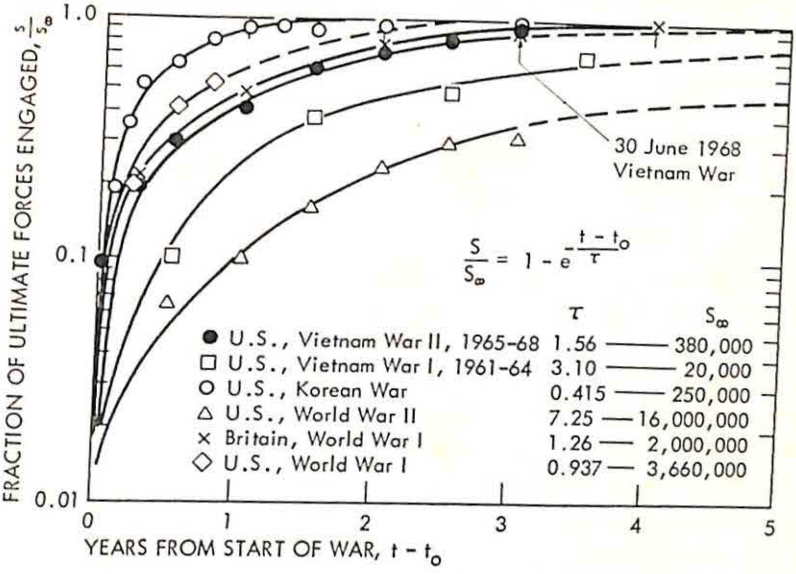
The constant, τ, which has the dimensions of time and is referred to as the time constant, characterizes the variation of troop strength in the battle zone, S, with respect to time. For the case of Vietnam War II, as shown in Figure 8b, one finds that $\tau = 1.56$ years; i.e., in the first 1.56 years, 63 percent of the troops that are to be committed to the battle zone will have been committed. In the next 1.56 years, 86 percent of the troops will have been committed; and in the next 1.56 years (yields a total of 4.68 years), 95 percent of the ultimate forces will have been committed. If the coupling relationships are assumed valid, then the enemy's strength in the battle zone will have increased by the same percentages.

In like manner, the battle strength to time relationship, plotted as the fraction of ultimate forces engaged in the battle zone, is shown in Figure 9a by a graphic representation of a summary of the United States statistics for the Vietnam War, the Korean War, World War II, and World War I (including the British Army). Figure 9a has been constructed in the same manner as Figure 8a in the light of the last equation. Here the strength of the United States Army in the battle zone has been chosen for the variable S. S_∞ for each war is estimated. For Vietnam Wars I and II, Korean, and British World War I, S_∞ can be estimated by inspection from a graph of strength *versus* time for each war in the same manner, as $S_\infty = 600{,}000$ is observed for Department of Defense strength in Vietnam War II in Figure 8a. The estimated value for the United States Army in these cases is as given in Figure 9a.

Not so obvious is the ultimate strength that could have been mobilized by the United States in World War I and World War II. In these two cases the wars ended before the curves became sufficiently horizontal to determine by inspection the value of S_∞. Thus, for World War II a value of 16 million is chosen for S_∞, because this was estimated to be the absolute ceiling on the number of men who could be considered physically fit for active war service at that time (7). Similarly, an S_∞ value of 3,670,000 was chosen for United States participation in World War I, because it was the same percentage of our population in France as Great Britain's two million soldiers in France was of her total population in 1918.

SUMMARY OF MAJOR U.S. WARS
– U.S. ARMY –

a. Normalized Strength vs Duration



b. Normalized Strength vs Normalized Duration

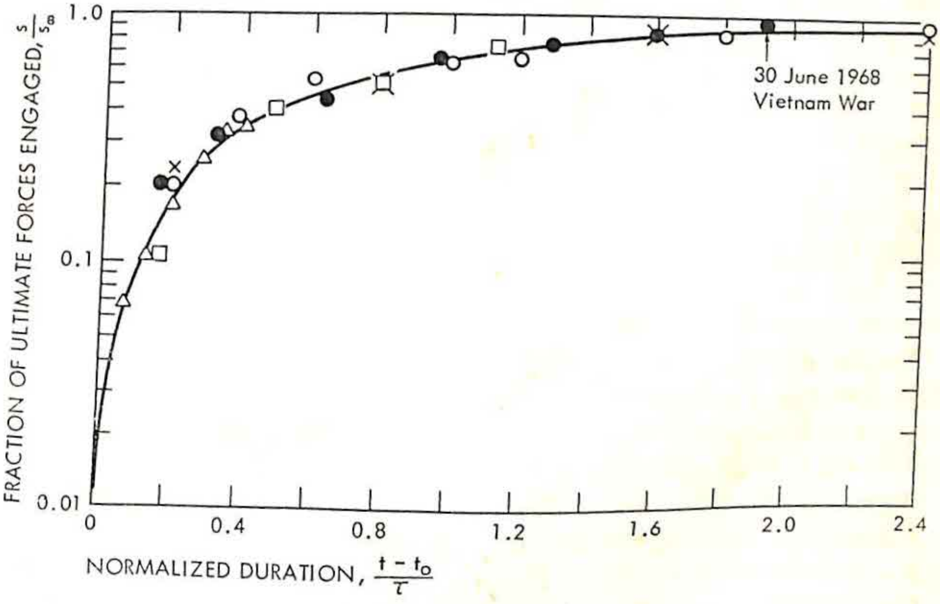


FIGURE 9
SUMMARY OF MAJOR U.S. WARS—U.S. ARMY

If one choose arbitrarily the initial time, t_0, such that S_0/S_∞ is at least less than .10, the variation of S/S_∞ for the different wars can then be plotted, and is shown in Figure 9a. The time constants for the different wars can then be determined.

The last equation shows that $S(t—t_0)/S_\infty$ is the same function of the parameter $(t—t_0)/\tau$ for all wars. This means that all wars can be brought together and represented by a single line when S/S_∞ is plotted against $(t—t_0)/\tau$ for each war. That this is true, is shown in Figure 9b for all wars analyzed.

This derivation and demonstration, in effect, shows that if the battle strength is expressed in terms of the percentage of ultimate forces to be engaged and is plotted against the time constant for each particular war, all wars studied can be expressed by a single mathematical relationship. Thus, each war forms a pattern over time and then history repeats each pattern.

Figure 9b allows comparison between the Vietnam War and all other wars analyzed. Where the Vietnam War stood, with respect to all other wars as of June 30, 1968, and every other six-month interval back to June 30, 1965, can be seen by the black data points. In sum it can be seen that 86 percent of the troops to be committed by both sides had been committed by June 30, 1968. It can also be seen that the rate of change from now on should stay minimal, providing those conditions that have been acting over the past three years continue to act unchanged in the future.

Deviations of the actual data points from the theoretical curves can be traced to a number of causes, such as the effects of elections, weather, and influenza epidemics. However, the major phenomenon seen here is the remarkable stability and trueness-to-course that, in general, wars follow over long periods of time. Instead of complete chaos, which is the expected characteristic of the conduct of human affairs, the exact opposite condition seems to exist when the human affair is warfare. When viewed in the manner demonstrated here, wars appear to become remarkably orderly events.

H. IMPLICATIONS FOR THE VIETNAM WAR

Because of the orderliness of our behavior, the implication of the Vietnam data is that we, the United States, our allies, and our enemies, as of this writing, are approaching a crisis point, where either a settlement is possible or another major escalation is indicated.

What prompts this statement is the shape of the curves for each of the previous wars and the shape of the curves for the Vietnam War. The shape of the curves for the past wars, and particularly the manner in which they level off, suggests that peace is negotiated when death, casualty, and strength exponentially approach the horizontal.

The overall leveling of these vital statistics can be seen in the British and German Armies during World War I, leading to a conclusion in November, 1918 (as shown in Figure 6) and for the Union and Confederate Armies, leading to a conclusion in April, 1865 (as shown in Figure 7). The same leveling occurred in the Korean War, Figure 10, and led to a conclusion in July, 1953. An identical pattern can be traced in the first phase of the Vietnam War, Figure 11, which ended December, 1964. It can be further noted that the Korean War and the Vietnam War are relatively close together in time and place; thus their comparability is best of all the other wars.

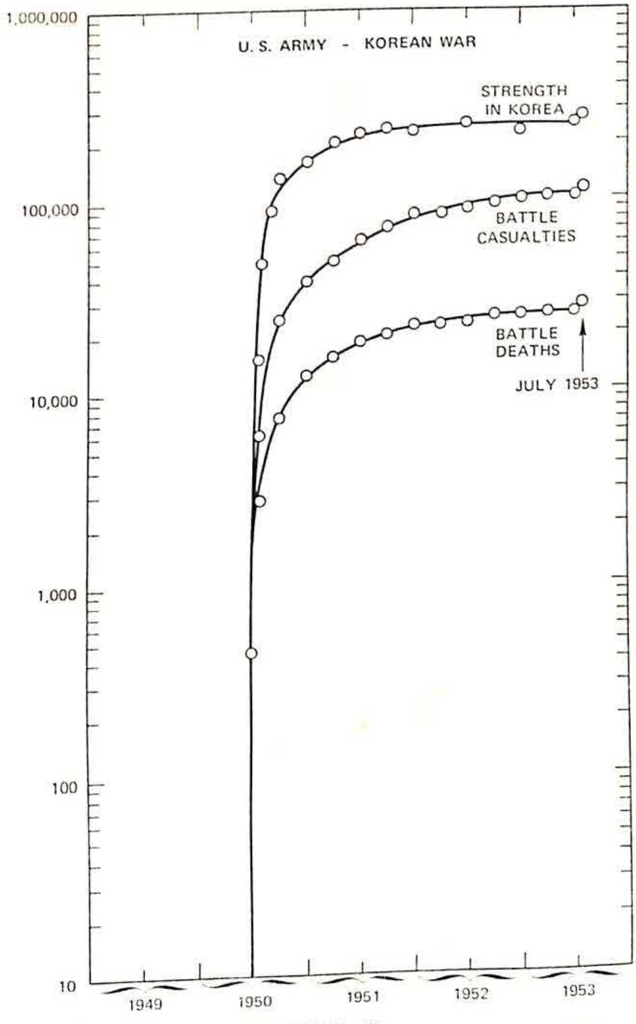


FIGURE 10
U.S. ARMY—KOREAN WAR

If these variables for the second phase of the Vietnam War continue to level in the latter half of 1968 and during the first half of 1969, then the past pattern will have repeated itself for the sixth time in our history.

As of this writing, since most of the principal nations of the world are supplying money, material, men, or some combination of the three, indeed, in reality, a global decision period is imminent.

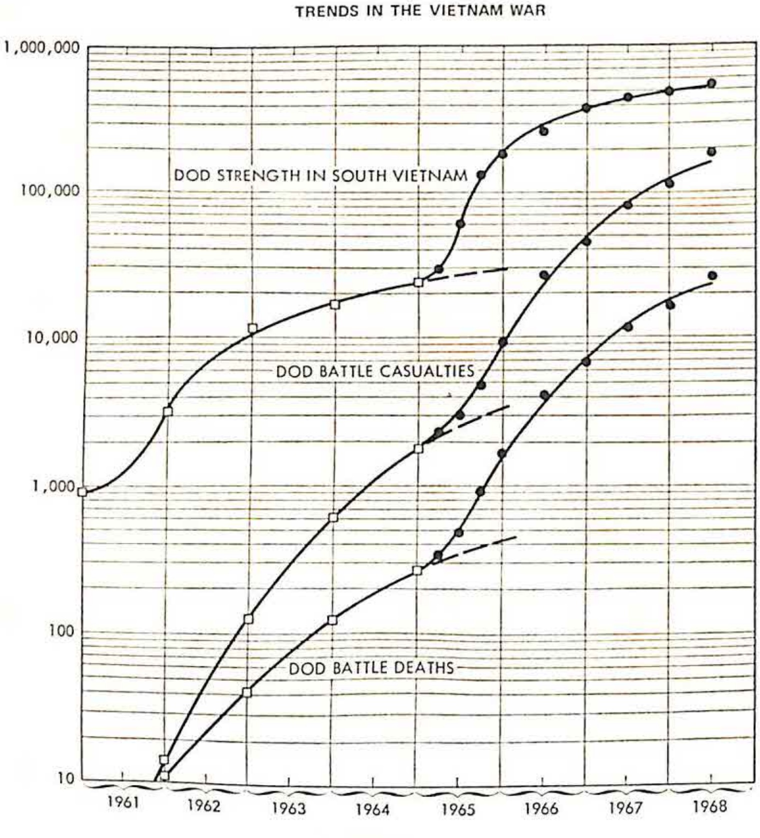


FIGURE 11
TRENDS IN THE VIETNAM WAR

I. SUMMARY

During the last 100 years, the United States has fought four major wars to conclusion: the Civil War, World War I, World War II, and the Korean War. For the past eight years, we have been engaged in a fifth and not yet concluded war, the war in Vietnam, which has occupied our minds, our troops, and our resources longer than any other war in our history. And yet, the Vietnam War appears to be progressing in the same orderly manner as our four previous wars, despite its record duration. The repetitive behavioral pattern suggests that we are approaching a crisis point in the Vietnam War where either a settlement is possible or another major escalation is indicated.

This analysis is based on a purely descriptive examination of the data—the vital statistics—of modern nations at war, particularly the United States, our allies, and our enemies during the last 100 years. No attempt is made to explain the processes which result in the behavioral orderliness. The data of all five wars are seen to fit very closely the same, simple, mathematical laws.[3]

REFERENCES

1. AYRES, L. P. The War with Germany—A Statistical Summary (1st ed.). Washington: Govt. Printing Office, 1919. P. 15.
2. CHURCHILL, W. S. The World Crisis (Vol. II). New York: Scribner, 1924. P. 39.
3. DEPARTMENT OF DEFENSE, OFFICE OF THE COMPTROLLER OF THE ARMY, PROGRAM REVIEW AND ANALYSIS DIVISION. ROK and UN ground forces strength in Korea, 31 July 1950—31 July 1953. Washington, D.C., 7 October 1964.
4. DEPARTMENT OF DEFENSE. Annual report of the Secretary of the Army—1948. Government Printing Office, Washington, D.C., 1949. P. 295.
5. LANCHESTER, F. W. Aircraft in Warfare; The Dawn of the Fourth Arm. London, England: Constable, 1916.
6. LIVERMORE, T. L. Numbers and Losses in the Civil War in America: 1861—1865. Bloomington, Ind.: Indiana Univ. Press, 1957. Pp. 75-141.
7. MARSHALL, G. C. Biannual Report to the Chief of Staff of the U.S. Army, July 1, 1943—June 30, 1945. Washington, D.C.: Govt. Printing Office, 1945. P. 101.
8. RICHARDSON, L. F. Arms and Insecurity. Pittsburgh, Pa.: Boxwood, 1960. Pp. 1-36.
9. STEVENS, S. S. (Ed.). Handbook of Experimental Psychology. New York: Wiley, 1966. Pp. 1-49.
10. VOEVODSKY, J. A mathematical representation of time-variant quantitative behavior. *J. of Psychol.,* 1968, **68**, 129-140.
11. VOEVODSKY, J., COOPER, L. M., MORGAN, A. H., & HILGARD, E. R. The measurement of suprathreshold pain. *Amer. J. Psychol.,* 1967, **80**, 124-128.
12. WIENER, N. Cybernetics. Cambridge, Mass.: M.I.T., 1962. Pp. 95-115.

ADDENDA

13. DEPARTMENT OF THE ARMY, OFFICE OF THE ADJUTANT GENERAL, STATISTICAL AND ACCOUNTING BRANCH, PROGRAM REVIEW AND ANALYSIS DIVISION OF THE OFFICE OF THE COMPTROLLER OF THE ARMY. Army battle casualties and nonbattle deaths in World War II, 7 December 1941-31 December 1946. OCS Report Control Symbol CSCAP (OT) 87. Pp. 10 & 95.
14. DEPARTMENT OF THE ARMY, OFFICE OF THE ASSISTANT CHIEF OF STAFF. Battle casualties of the Army, 30 September 1954. CSGPA-363.
15. DEPARTMENT OF DEFENSE, OFFICE OF THE ASSISTANT SECRETARY OF DEFENSE (PUBLIC AFFAIRS). Vietnam weekly casualties, statistical summary, December 1961-June 1968.

16. THE WAR OFFICE. Statistics of the military effort of the British Empire during the Great War, 1914-1920. HM Stationery Office, London, England, 1922. Pp. 358-362.

36 El Rey Road
Portola Valley, California 94025

---

[3] For World War II (3), all theaters and campaigns, including the Army Air Corps officers and enlisted men, are grouped in U.S. Army casualties, except from 7 December 1941 to 10 May 1942. During this period three noncombat divisions were captured and held for the duration of the war. This resulted in 30,838 casualties and 13,847 deaths being listed as having occurred in the single month of May, 1942. This accounting procedure thus caused the only significant statistical anomaly noted in all the data. Hence these data were excluded. For the U.S. participation in World War I (1, pp. 120-122), battle deaths include only deaths from action. They do not include deaths from diseases and privation. Battle deaths for the month of May, 1917, were estimated as 400, the month of April as 100, and the month of December as 800. With these estimates, a 4.4 percent difference exists between total battle deaths of 48,909 reported on page 122 and the accumulated total deaths on page 120. Total battle casualties at the end of the war were 286,330, of which 7347 were classified as missing in action or taken prisoner. 230,074 were the total wounded. The rest were killed. For certain U.S. Civil War battles (9, pp. 140-1), Confederate strength on the battlefield for a particular battle (not battle strength) had to be estimated and was assumed to be one-half of the known Union strength. This, in general, proved to be the case throughout the war. Estimates of Confederate battle casualties, when their derivation was necessary, were made by multiplying the Confederate battlefield strength by .123, which is the average Confederate casualty per soldier engaged in 30 battles of which 12 were defeats and 18 were victories. Estimates of Confederate battle deaths, when made necessary, were derived by multiplying Confederate battle casualties for that battle by 17.6 percent, which is the average for 29 battles. Estimates of Union battle deaths were made by multiplying Union battle casualties by 17.5 percent, which is the average for 37 battles. Union and Confederate missing and prisoners were not available.